What is beneath the surface? Option pricing with multifrequency latent states

Calvet, Laurent E ; Fearnley, Marcus ; Fisher, Adlai J ; Leippold, Markus

Abstract: We introduce a tractable class of non-affine price processes with multifrequency stochastic volatility and jumps. The specications require few xed parameters and deliver fast option pricing. One key ingredient is a tight link between jumps and volatility regimes, as asset pricing theory suggests. Empirically, the model matches implied volatility surfaces and their dynamics without requiring parameter recalibration. A variety of metrics show improvements over traditional benchmarks in- and out-of-sample.

DOI: https://doi.org/10.1016/j.jeconom.2015.02.034

What's Beneath the Surface ? Option Pricing with Multifrequency Latent States

Laurent E. CALVET
HEC School of Management, GREGHEC, Paris, France

Marcus FEARNLEY
Department of Finance, HEC School of Management, Paris, France

Adlai FISHER
Sauder School of Business, University of British Columbia, Vancouver, Canada

Markus LEIPPOLD
University of Zurich, Zurich, Switzerland

CR 969-2013
ISBN : 2-85418-969-8

ISBN : 2-85418-969-8

What's Beneath the Surface? Option Pricing with Multifrequency Latent States

Laurent E. Calvet, Marcus Fearnley, Adlai J. Fisher, and Markus Leippold*

January 18, 2013

Abstract

We introduce a tractable class of non-affine price processes with multifrequency stochastic volatility and jumps. The specifications require few fixed parameters and deliver fast option pricing. One key ingredient is a tight link between jumps and volatility regimes, as asset pricing theory suggests. Empirically, the model matches implied volatility surfaces and their dynamics without requiring parameter recalibration. A variety of metrics show improvements over traditional benchmarks in- and out-of-sample.

Keywords: Markov-switching multifractal, particle filter, regime-switching, stochastic volatility, jump-risk premium, option pricing.

JEL Classification: C51, G12, G13.

*Calvet: Department of Finance, HEC Paris, 1 rue de la Libération, 78351 Jouy en Josas, France; calvet@hec.fr. Fearnley: Department of Finance, HEC Paris, 1 rue de la Libération, 78351 Jouy en Josas, France; marcus.fearnley@hec.edu. Fisher: Sauder School of Business, University of British Columbia, 2053 Main Mall, Vancouver, BC Canada V6T 1Z2; adlai.fisher@sauder.ubc.ca. Leippold: University of Zurich - Department of Banking and Finance, Plattenstrasse 14, 8032 Zurich, Switzerland; markus.leippold@bf.uzh.ch. We received helpful comments from John Campbell, René Garcia, Rodney Hoskinson, Ioanid Rosu, Ken Singleton, and seminar participants at HEC Paris and the University of Zurich.

1 Introduction

A variety of statistical models of returns have been used to fit option implied volatility (IV) surfaces at a given date. Under correct specification of objective and risk-neutral densities, such pricing models should be dynamically consistent, and the need to frequently recalibrate parameters should be modest. Despite great progress over the past thirty years, a central challenge still remains in the option pricing literature: to develop parsimonious price processes characterized by a small number of stable, well-identified parameters, yet with rich enough dynamics to match greatly varying shapes of the IV surface at different dates.

Latent factor models seem well-suited to this challenge since changes in hidden states can drive variation in conditional return densities, while a small set of well-identified parameters remains fixed over time. Many parametric latent factor specifications have emerged over the last two decades, with canonical examples being the stochastic volatility models of Hull and White (1987) and Heston (1993) and the affine jump-diffusions proposed by Bates (1996), Bakshi et al. (1997), and Duffie et al. (2000). Qualitatively, these specifications capture key statistical features including variance autocorrelation and return asymmetry. Nonetheless, considerable evidence has accumulated showing that one-factor affine models are too restrictive,[1] motivating recent research into multifactor and non-affine specifications.

Consistent with the challenges faced by single-factor affine specifications, evidence has gathered in favor of permitting multiple components for volatility modeling (Calvet et al. (1997), Engle and Lee (1999), Calvet and Fisher (2001, 2008a)) and volatility forecasting (Calvet and Fisher (2004, 2006), Lux (2008), Corsi (2009)).[2] In the empirical option pricing literature, multiple components have been incorporated into both GARCH-type and stochastic volatility models, and both approaches have shown that short-run and long-run components are helpful to capture the term structure of the option implied volatility surface. Component GARCH-type models, as in Christoffersen et al. (2008), are easier to estimate since the volatility state is captured fully by past returns.[3] However,

[1] See, for example, Garcia et al. (2009), Jones (2003), Li and Zhang (2013), and the literature cited therein.

[2] See also Adrian and Rosenberg (2008), Bates (2000, 2012), Campbell et al. (2012), Chernov et al. (2003), Christoffersen et al. (2009), Christoffersen, Dorion, Jacobs and Wang (2010), Egloff et al. (2010), and Lux and Kaizoji (2007).

[3] See also Corsi et al. (2013), who incorporate information from realized volatility in a three-component specification.

economic intuition suggests that the option surface, which reflects forward-looking investor information, should contain a great deal of information about important state variables beyond the history of returns. Stochastic volatility models with multiple latent components offer the possibility of capturing information in the option surface, but are much more challenging to implement empirically. Bates (2000) develops an estimation method based on least-squares fitting of option prices and offers an early analysis of an affine jump-diffusion with two volatility factors. He concludes that jumps are necessary to provide a plausible match between objective and risk-neutral densities. Christoffersen et al. (2009) show the importance of incorporating a second volatility component for option prices using an estimation method similar to Bates (2000), without incorporating jumps. Andersen et al. (2012) develop a method for incorporating realized volatility into estimation, and consider extensions of the two-component stochastic volatility model with jumps. A variety of non-affine approaches have also been considered. These permit additional flexibility, but typically face important challenges in empirical implementation.[4] In all of this literature, frontier-level research topics include the parsimonious specification of multi-factor stochastic volatility models with jumps as well as the efficient extraction of information from the joint density of returns and option prices.

To contribute to this agenda, we develop a class of non-affine yet highly tractable models. Our approach allows parsimonious specification of both objective and risk-neutral densities, while at the same time offering a rich state space, driving multifrequency stochastic volatility and jumps, that helps to match the variety of option surfaces seen in the data. We also contribute methodologically by adapting particle filtering methods (e.g., Johannes et al. (2009)) to permit efficient estimation of option pricing parameters using both returns and option surfaces.

Our model uses a base volatility specification that parsimoniously accommodates multiple volatility components of heterogeneous frequencies, following the Markov-switching Multifractal ("MSM") approach of Calvet and Fisher (2001) and Calvet and Fisher (2004). MSM is a pure regime-switching model that matches commonly observed features of asset returns such as a hyperbolic autocorrelation in volatility, power-law scaling of return moments, and fat tails of the return distribution. This

[4]Because closed-form solutions are typically unavailable, empirical studies of non-affine models are most often based either on pure time-series evidence or are restricted to diagnosing simple volatility contracts such as variance swaps or the VIX index. In recent literature, Kaeck and Alexander (2012) estimate a variety of affine and non-affine models on S&P 500 index returns and the VIX term structure. Christoffersen, Jacobs and Mimouni (2010) advance the estimation of single-factor non-affine models using a particle filter, as discussed below.

base model provides a rich state space that has been shown to perform well in-and-out of sample in volatility forecasting and conditional density estimation relative to standard benchmarks.

The present paper extends MSM using key ingredients from prior literature, producing a model that we call Skew MSM. First, we add a diffusion component to stochastic volatility. Following Heston (1993), the diffusion component is correlated with innovations to returns, accommodating negative volatility-return correlation, also known as the "leverage" effect. We provide empirical evidence to show that adding this diffusion component is helpful. In particular, while volatility itself decays at a slow hyperbolic rate, the return-volatility autocorrelation has a component that decays very quickly, which we capture by permitting a separate decay rate for innovations to the diffusive volatility component. Skew MSM nests both standard MSM and the basic Heston (1993) specification, and we are able to show the importance of each of these building blocks in fitting the joint dynamics of the underlying index and option prices.

An additional key ingredient of Skew MSM consists of jumps in asset prices that are linked to regime changes in volatility, following from the implications of the equilibrium model of Calvet and Fisher (2008b). Prior literature such as Duffie et al. (2000) emphasizes the potential importance of jumps in matching essential features of stock returns and option prices.[5] At the same time, it is well known that empirical identification of jump parameters, especially for the rarest and therefore largest events, is difficult. We overcome this problem by linking price jumps to changes in volatility as suggested by equilibrium theory, and by the strong but empirically valid structure placed on volatility changes by the base MSM model. The resulting jump structure in Skew MSM yields price discontinuities of heterogeneous size and frequency, with many small jumps and a few large jumps.

The final key ingredient of Skew MSM is the specification of risk premia. Also following from equilibrium theory, the structure we place on risk premia permits that under the risk-neutral density, investors price assets as if increases in volatility are more likely than their probability under the objective density, and as if decreases in volatility are less likely than under the objective density. This feature strengthens the unconditional skewness of the underlying asset under the risk-neutral density, which is important to explaining the pervasive "smirk" observed in implied volatilities since

[5]See also Bates (1996, 2000), Broadie et al. (2007), Carr et al. (2002), Carr and Wu (2004), Liu et al. (2005), Merton (1976), Naik and Lee (1990), and Pan (2002).

risk-neutral implied skewness tends to be much stronger than unconditional skewness under the objective density.

The class of processes we arrive at is non-affine because price dynamics are subject to abrupt changes in regime. Option pricing using regime switching has been the subject of considerable study in the mathematical finance literature because of the rich dynamics permitted by this class of models (e.g., Elliott, Siu and Chan (2007), Elliott, Siu, Chan and Lau (2007)).[6] However, the pricing of general regime-switching models requires the computationally expensive solution of a system of partial differential equations. As a consequence prior empirical research on option pricing using regime-switching frameworks has been limited.[7] By contrast, Skew MSM is conditionally affine and the pricing equations reduce to a system of ordinary differential equations, permitting fast option valuation. We thus depart from the affine class, but our conditionally affine approach remains highly tractable for the types of empirical option pricing applications that are common in industry.

We develop filtering and estimation techniques for Skew MSM. The latent state can be tracked over time by way of a particle filter, a recursive algorithm that uses Monte Carlo draws (particles) to approximate the conditional state and data densities at each time-step.[8] The original sampling and importance resampling approach to particle filtering, which was introduced by Gordon et al. (1993), has recently been used in volatility filtering applications by Calvet et al. (2006), Johannes et al. (2009), Malik and Pitt (2011), and Christoffersen, Jacobs and Mimouni (2010). An issue that arises in a model with a rich state space and large rare events is that random sampling of particles may not capture sudden changes into unlikely, but nonetheless important, parts of the state space. To confront this issue we develop a variant of the standard particle filter by incorporating a stratified sampling approach, similar in spirit to Kitagawa (1996), which ensures that all parts of the discrete MSM state space are represented.

The particle filter permits maximum likelihood estimation (MLE) of Skew MSM, which is based

[6] See also Elliott et al. (2011), which studies a restricted version of the equilibrium model developed in Calvet and Fisher (2008b).

[7] In a recent contribution, Durham and Park (2012) circumvent the problem of using the cross-section of option prices by estimating their model on integrated volatility.

[8] Alternatives include the approximate maximum likelihood approach, applied to filtering and estimation of affine latent factor processes by Bates (2006, 2012). In earlier studies, method of moments estimation is common (Andersen and Sorensen (1996), Pan (2002), Chernov and Ghysels (2000)).

by construction on the joint density of the underlying asset returns and a panel of option prices. In comparison, Johannes et al. (2009) estimate jump-diffusions using stock return data only, while incorporating short-maturity options into filters to generate diagnostics. Christoffersen, Jacobs and Mimouni (2010) use option data to evaluate goodness of fit and estimate parameters, but use only stock returns, and do not incorporate option prices, into their filter for states. To the best of our knowledge, our paper is the first in the literature to carry out efficient estimation of parameters and filtering of latent states using the joint density of returns and option surfaces, even for benchmark single-factor affine jump diffusions. Moreover, because of the tractability of option prices in our setting, we are also able to carry out filtering and MLE using the joint density of returns and option surfaces for our multi-component non-affine jump-diffusion.

Our empirical analysis begins by examining the performance of Skew MSM using only equity index returns. We conduct ML estimation on a long sample of the S&P 500 index, and find that the in-sample fit increases as frequency components are added, even though this requires no additional parameters. We show empirically that the diffusive volatility component helps to capture quick decay in the cross-correlation between returns and volatility (leverage effect), while the MSM components capture slow decay in volatility autocorrelations. A variety of benchmark comparison models, including standard jump diffusion and GARCH-type models, are not able to capture this empirical fact. Overall, Skew MSM shows considerably better in-sample fit on the returns data.

We then carry out joint estimation using both index return and option data. We make one concession to computational simplification, which is to use a two-step estimation procedure. In the first step, we estimate the objective density parameters on the time-series of equity returns only. In the second step, we use all available data, including stock returns and a large options panel, to estimate the remaining risk-premium parameters. Our in-sample options data is composed of almost ten years of monthly options surfaces, with a wide range of strikes and maturities. We find that Skew MSM fits the option surfaces in sample substantially better than the traditional affine jump-diffusions. Using an out-of-sample panel of an additional three years of data, Skew MSM again matches option prices much more closely.

These findings suggest that the ability to parsimoniously incorporate a high-dimensional state

space provides a useful advantage in modeling the joint behavior of stock and option prices. This conclusion mirrors, in a different setting, the findings of Calvet et al. (2011). This prior work shows that by parsimoniously specifying high-dimensional latent state models, interest rate term structures can be fit in a dynamically consistent manner with extremely high precision. The present paper demonstrates the fruitfulness of this approach for option pricing.

The remainder of the paper is structured as follows. In Section 2, we introduce Skew MSM. Section 3 shows how to efficiently price European options. In Section 4, we construct two particle filter methods for filtering and estimating Skew MSM. We also introduce benchmark jump diffusion and GARCH models. Section 5 contains our empirical analysis. Section 6 concludes.

2 The Skew MSM model of stock prices

In this section, we introduce a new class of regime-switching diffusions that incorporate the jumps, stochastic volatility, multiple frequencies and leverage exhibited by financial series.

2.1 A diffusion with regime-switch dependent jumps

We consider a frictionless financial market defined on the continuous time domain $\mathcal{T} = [0, \infty)$. The structure of uncertainty is specified by a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathcal{T}}, \mathbb{P})$, where $\mathbb{P}$ denotes the physical measure. The usual assumptions of right-continuity and completeness with respect to the null sets of $\mathbb{P}$ are assumed to hold.

Investors can trade a riskless asset with continuously compounded interest rate r_t. They can also trade a stock with price S_t and continuous dividend yield d_t. The excess return on the stock between dates t and $t + dt$ is correspondingly

$$\frac{dS_t}{S_t} + (d_t - r_t)dt. \tag{1}$$

The stock price S_t is driven by a Markov chain $\{M_t\}_{t\in\mathcal{T}}$ taking values on a finite set $\mathcal{D} = \{m^1, ..., m^d\}$. The continuous-time dynamics of the Markov chain are described by the transition rate matrix $Q = (q_{ij})_{1\leq i,j\leq d}$, which has the usual properties that $q_{ij} \geq 0$ for all $i \neq j$ and $q_{ii} = -\sum_{i\neq j} q_{ij}$ for all

i. The off-diagonal component q_{ij} quantifies the probability of switching from state i to state $j \neq i$ between t and $t + dt$:

$$\mathbb{P}\left(M_{t+dt} = m^j \middle| M_t = m^i\right) = q_{ij}dt + o\left(dt\right). \tag{2}$$

By contrast, the diagonal element q_{ii} controls the probability that the Markov chain exits state i:

$$\mathbb{P}(M_{t+dt} = m^i | M_t = m^i) = 1 + q_{ii}dt + o\left(dt\right)$$

between t and $t + dt$. For every $i, j \in \{1, \ldots, d\}$, $i \neq j$, we denote by $N_{i,j,t}$ the number of transitions from state i to state j between dates 0 and t. To simplify notation in the rest of the paper, we let $N_{i,i,t} = 0$ for all i and t.

We assume that under the statistical measure $\mathbb{P}$, the log-price process $s_t \equiv \log S_t$ evolves according to the system of stochastic differential equations:

$$ds_t = \mu_t dt + \sqrt{V_t}dW_{1,t} + dJ_t - \bar{J}_{M_t}dt, \tag{3}$$

$$dV_t = \kappa\left(\theta_{M_t} - V_t\right)dt + \sigma\sqrt{V_t}\left(\rho dW_{1,t} + \sqrt{1-\rho^2}dW_{2,t}\right), \tag{4}$$

where $W_t = (W_{1,t}, W_{2,t})'$ is a bivariate Wiener process, and $\kappa, \sigma \in \mathbb{R}_{++}$ and $\rho \in (-1, 1)$ are fixed parameters.[9] Equation (3) implies that infinitesimal price changes are driven by: (i) a stochastic variance process V_t; (ii) a compensated jump process triggered by changes in the Markov chain, $dJ_t - \bar{J}_{M_t}dt$; and (iii) a time-varying drift μ_t. We now discuss these three building blocks of the log-price specification.

By (4), stochastic variance is a diffusive process that mean-reverts toward θ_{M_t}, where $\theta_{M_t} \in \mathbb{R}_+$ is a deterministic function of the Markov state M_t. As in Heston (1993), the Gaussian innovation to V_t is correlated with the Gaussian innovation to the stock price, which accommodates high-frequency leverage effects. We extend the Heston model by allowing regime-switching in "long-run" level θ_{M_t}. This innovation allows us to incorporate lower-frequency fluctuations into the stochastic variance V_t,

[9] The switching points of the Markov chain partition the time domain $\mathcal{T} = [0, \infty) = \bigcup_{i=0}^{\infty}[t_i, t_{i+1})$ into subintervals over which the drift θ_{M_t} is constant and a strong solution V_t to (4) is known to exist (e.g, Øksendal (2010)). The variance process V_t is therefore well-defined on the entire domain. Furthermore, V_t is strictly positive if it satisfies the Feller condition $2\kappa\theta_{m^i} \geq \sigma^2$ for all $i \in \{1, \ldots, d\}$. The Feller condition tends to be binding in the estimation of a simple Heston (1993) model on equity returns. By contrast, we have found that the Feller conditions are not binding in the estimation of the regime switch-dependent jump-diffusions considered in Section 5. This property is another advantage of our approach, since violations of the Feller conditions are generally associated with numerical instabilities (e.g., Lord et al. (2010)).

while preserving the tractability of the Heston model.

We assume that the jump process is triggered by changes in the Markov state M_t, which tightly knits the dynamics of returns and volatility. Specifically, we consider a set of fixed coefficients $J_{i,j}$, $i, j \in \{1, \ldots, d\}$, and define the process J_t as

$$J_t = \sum_{1 \leq i,j \leq d} J_{i,j} N_{i,j,t}. \tag{5}$$

Since by definition $N_{i,i,t} = 0$, the diagonal coefficients J_{ii} do not contribute to the definition of J_t, and we assume without loss of generality that they are equal to zero. By (3) and (5), the log price of the stock jumps by $J_{i,j}$ when the Markov chain switches from state i to state j. The process J_t is therefore the cumulative jump between dates 0 and t. By Ito's lemma, the excess return on the stock (1) is equal to

$$\left(\mu_t + d_t + \frac{V_t}{2} - r_t\right) dt + \sqrt{V_t} dW_{1,t} + e^{dJ_t} - 1 - \bar{J}_{M_t} dt.$$

We select $\bar{J}_{M_t}$ so that the jump process makes no contribution to the expected excess return. This implies that

$$\bar{J}_{M_t} = \sum_{j=1}^{d} q_{i,j}(e^{J_{i,j}} - 1) \tag{6}$$

if the Markov chain is currently in state $M_t = m^i$. Our model nests Heston (1993)'s specification for s_t and V_t if the function θ is constant and there are no jumps.

We specify the time-varying drift μ_t as

$$\mu_t = r_t - d_t + \left(\alpha - \frac{1}{2}\right) V_t, \tag{7}$$

where α is a constant. The specification implies that the conditional expected excess return on the stock is proportional to the variance V_t:

$$\mathbb{E}^{\mathbb{P}}\left[\frac{dS_t}{S_t} + (d_t - r_t)dt \middle| M_t, V_t\right] = \alpha V_t dt,$$

as is the case in the consumption CAPM and its many extensions (e.g., Merton (1980)).

We will verify in Section 3 that the log-price s_t is conditionally affine and that its characteristic function can be obtained by solving a system of ordinary differential equations. These properties will allow us to price options quickly using transform-based methods.

2.2 A multifrequency specification: Skew MSM

We now provide a specialized version of the jump-diffusion that is both parsimonious and sufficiently rich to capture the multifrequency dynamics of equity returns.

Markov chain. In order to accommodate the rich volatility features of equity series, we assume that the Markov chain M_t follows a Markov-switching multifractal process (Calvet and Fisher (2001), Calvet and Fisher (2008a)), as we now explain. The state M_t is a vector with $\bar{k}$ components:

$$M_t = \left(M_{1,t}, M_{2,t}, ..., M_{\bar{k},t}\right) \in \mathbb{R}_+^{\bar{k}}. \tag{8}$$

The components of the state vector are mutually independent across k, and each component evolves through time at its own frequency. The construction is also based on a fixed distribution M that has a unit mean and a positive support.

The state vector M_t evolves dynamically as follows. Over an incremental period of time dt, each component $M_{k,t}$ is either drawn from the distribution M with probability $\lambda_k dt$, or remains constant with probability $1 - \lambda_k dt$. The dynamics of each component can be summarized as follows:

$$\begin{array}{ll} M_{k,t+dt} \text{ drawn from the distribution } M & \text{with probability } \lambda_k dt \ , \\ M_{k,t+dt} = M_{k,t} & \text{with probability } 1 - \lambda_k dt \ . \end{array}$$

For parsimony, we assume that intensities follow a geometric progression:

$$\lambda_k = \lambda_1 b^{k-1} \tag{9}$$

for every $k \in \{1, \ldots, \bar{k}\}$. As in Calvet and Fisher (2004), we choose M to be a binomial distribution, which can take the high value $m_0 \in [1, 2)$ or the low value $2 - m_0 \in (0, 1]$ with equal probability. The Markov chain M_t can therefore take $d = 2^{\bar{k}}$ possible values $m^1, ..., m^d \in \mathbb{R}_+^{\bar{k}}$, and the rate matrix Q can easily be written as a function of the intensities $\lambda_1, ..., \lambda_{\bar{k}}$.

Variance process. The "long-run" level of the variance V_t is defined as the rescaled product of the state components:

$$\theta_{M_t} = \bar{\theta} \prod_{k=1}^{\bar{k}} M_{k,t}. \tag{10}$$

By construction, θ_{M_t} follows a Markov-switching multifractal variance process, as defined in Calvet

and Fisher (2001, 2004, 2008a). Since M_t has a finite support, the long-run process θ_{M_t} is discrete as well.

The diffusive variance process V_t defined by (4) accomplishes two main purposes. First, V_t is a diffusive extension of MSM that can take any positive real value. This property provides sufficient flexibility to match the option surface. Second, the variance process can capture the empirically observed negative correlation between returns and innovations to volatility. This cross-correlation is often referred to as the "leverage" effect. The MSM process θ_{M_t} ensures that variance is highly persistent and (unconditionally) non-affine. The combination of both θ_{M_t} and V_t allows the model to capture the slowly decaying autocorrelation in volatility and the fast decay in the leverage effect exhibited by the data.

Jumps. We specify the deterministic jumps as intensity-weighted sums of the signs of component changes:

$$J_{ij} = \beta \sum_{k=1}^{\bar{k}} \frac{1}{\lambda_k} \text{sgn}\left(m_k^j - m_k^i\right) \tag{11}$$

where m_k^i denotes the kth component of state $m^i \in \mathbb{R}^{\bar{k}}$ for all i, k, and the sign function $\text{sgn}(x)$ equals -1 if x is strictly negative, 0 if x equals zero, and $+1$ if x is strictly positive. The intuition for imposing this structure is that when an MSM volatility component switches from a low to a high value (bad news), there will be a negative shock to returns as the investor prices in the higher expected future volatility. Similarly, a switch from a high volatility state to a low state (good news) will result in a large positive shock. The magnitude of the shock increases monotonically with the persistence of the component that switches.

The deterministic jump process of Skew MSM is inspired by the equilibrium model of Calvet and Fisher (2008b), in which price jumps are endogenously generated when an MSM dividend process switches between Markov states. Intuitively, when investors learn about changes in long-term volatility, they immediately factor this into the price of the asset, producing a discrete jump. Furthermore, the chosen specification (11) has the property that the price jump triggered by a transition from state j to state i is the negative of the price jump triggered by a transition from i to j: $J_{j,i} = -J_{i,j}$, as is the case in the equilibrium model of Calvet and Fisher (2008b). Skew MSM therefore differs from

the more common compound Poisson jumps, and in particular imposes a theoretically-motivated dependence between the volatility and jump processes.

The Skew MSM specification. The resulting model of the stock price under the physical measure $\mathbb{P}$, which we call Skew MSM, is fully specified by the nine parameters:

$$\Psi_{\mathbb{P}} \equiv \left(m_0, \lambda_{\bar{k}}, b, \bar{\theta}, \kappa, \sigma, \rho, \beta, \alpha\right), \tag{12}$$

where m_0 controls the levels of the Markov chain M_t, $\lambda_{\bar{k}}$ and b control its transition rates, $\bar{\theta}$ is the long-run level of the stochastic variance V_t, κ is its adjustment speed, σ controls its volatility, ρ denotes the correlation of the Gaussian innovations driving the log-price and stochastic variance, β controls the magnitude of the jumps, and α controls the sensitivity of the price drift to the stochastic variance. Skew MSM reduces to the Heston (1993) model if $m_0 = 1$ and $\beta = 0$, and to binomial MSM if $\kappa = +\infty$ and $\sigma = 0$.

Figure 1 illustrates the various components of the Skew MSM model. The model retains key statistical features of both the Heston and MSM models and has a unique jump structure that differs from the compound Poisson jumps that are typically used in the jump-diffusion literature.

The behavior of Skew MSM as the number of components gets large is of particular importance for empirical applications. For instance, it is informative about the marginal benefits of increasing the state space and helps guide the selection of $\bar{k}$. MSM is known to converge to a limiting process as the number of components $\bar{k}$ goes to infinity, provided that the condition $\mathbb{E}(M^2) < b$ holds (Calvet and Fisher (2001, 2008a)). Skew MSM is a diffusive extension of MSM, and we now show that it also converges to a square-integrable limiting process as $\bar{k} \to \infty$. Consider a bounded interval $[0, T]$, and let $L^2(\Omega \times [0,T])$ denote the Hilbert space of adapted square-integrable processes $y : \Omega \times [0,T] \to \mathbb{R}$. We also consider the norm:

$$\|y\|_{L^2(\Omega\times[0,T])} = \left[\mathbb{E}\left(\int_0^T y_t^2 dt\right)\right]^{1/2}.$$

We show in Appendix A:

Proposition 1 (Convergence to a limiting process as $\bar{k} \to \infty$). *Consider a fixed set of parameters m_0, λ_1, b, $\bar{\theta}$, κ, σ, ρ, β, and α, and a fixed set of initial values s_0 and V_0. Let $\{M_{k,t}\}_{k=1}^{\infty}$*

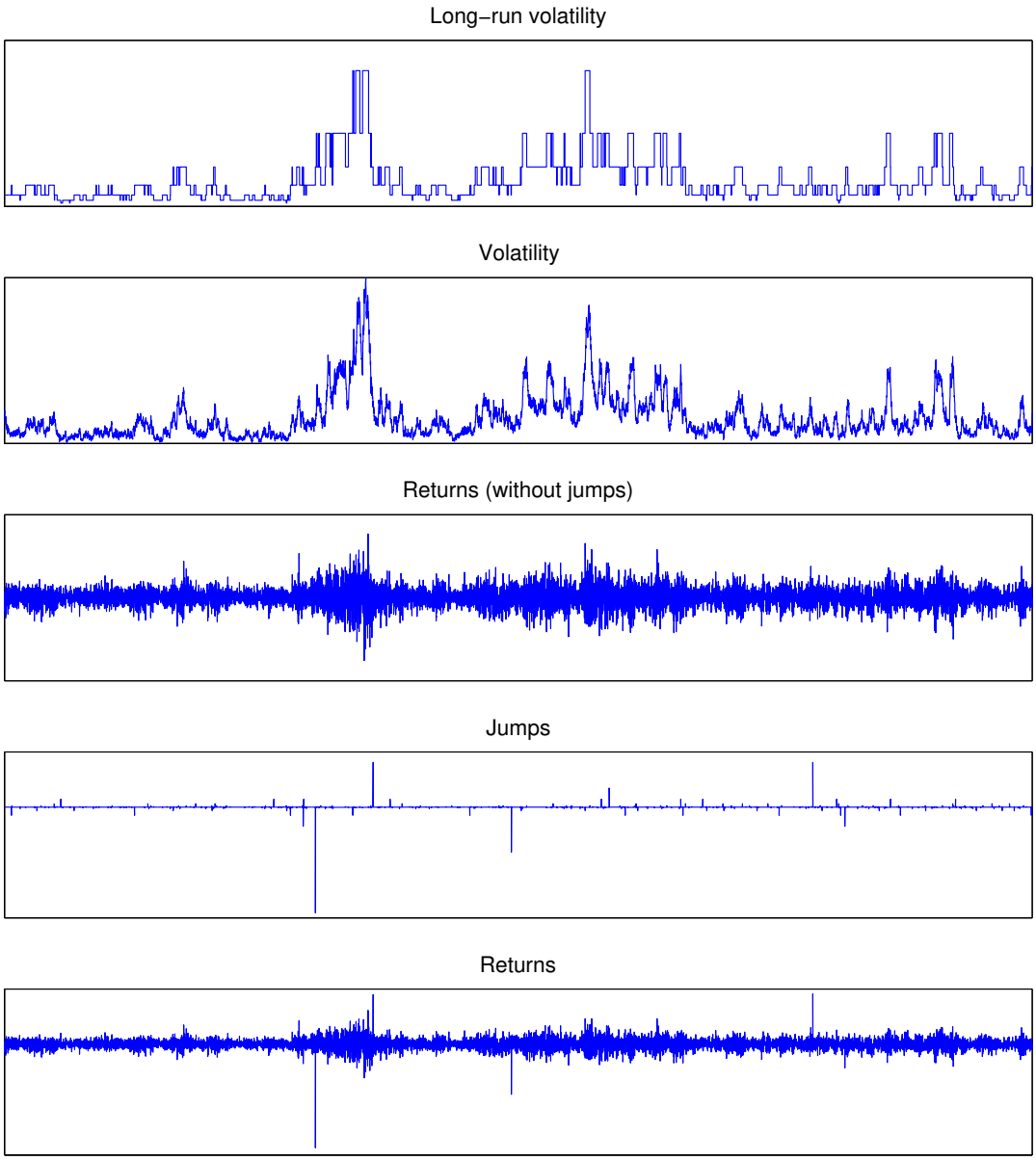

Figure 1: Simulation of Skew MSM process. The figure illustrates the time-series behavior of the various Skew MSM stochastic components using simulated data. From top to bottom: the MSM process θ_{M_t}; the volatility process V_t; the diffusive return component; the return jump component; and the combined log-return process.

denote a sequence of MSM Markov state variables with intensities $\lambda_k = \lambda_1 b^{k-1}$. *If* $\mathbb{E}(M^2) < b$, *the stochastic variance* V_t *and log-price* s_t *converge to limiting processes in* $L^2(\Omega \times [0,T])$ *as the number of components* $\bar{k}$ *goes to infinity.*

The convergence property implies that incorporating new factors will have only marginal effects on the stock return process when $\bar{k}$ is large. This property is important for selecting the number of factors in empirical work, as will be discussed in Section 5.

2.3 The market prices of risk

The latent volatility and regime-switches in our model render the market incomplete with respect to the risk-free rate and underlying stock price. For this reason, we assume that the risk-neutral measure $\mathbb{Q}$ has Radon-Nikodym derivative

$$\left.\frac{d\mathbb{Q}}{d\mathbb{P}}\right|_t = \xi_t^{(1)}\xi_t^{(2)}$$

with respect to the physical measure $\mathbb{P}$, where $\xi_t^{(1)}$ modifies the dynamics of the Markov chain M_t and $\xi_t^{(2)}$ modifies the drift of the Wiener process W_t and the compensator of the jump process. The exact definitions of $\xi_t^{(1)}$ and $\xi_t^{(2)}$ are provided in Appendix B.

Under the risk-adjusted measure $\mathbb{Q}$, the state M_t is a Markov chain with transition rates:

$$\begin{aligned} q_{ij}^* &= q_{ij}\Lambda_j/\Lambda_i \qquad \text{if } i \neq j, \\ q_{ii}^* &= -\textstyle\sum_{k=1,k\neq i}^{d} q_{ik}^* \quad \text{if } i = j. \end{aligned} \tag{13}$$

We denote by $Q^* = (q_{ij}^*)_{1\leq i,j\leq d}$ the corresponding transition rate matrix. Intuitively if state j is riskier than state i, then state j a has higher risk premium ($\Lambda_j > \Lambda_i$); specification (13) implies that a switch from the good state i to the bad state j is then more likely under $\mathbb{Q}$ than under $\mathbb{P}$. To retain the parsimony of the MSM process, we assume that the risk premia associated with each Markov state are:

$$\Lambda_i = \prod_{k=1}^{\bar{k}} \left[(b^*)^{\bar{k}-k}\ 1_{\{m_k^i = 2-m_0\}} + \frac{1}{(b^*)^{\bar{k}-k}}\, 1_{\{m_k^i = m_0\}} \right], \tag{14}$$

where $b^* \in [1,\infty)$ is fixed. Under this parametrization, the kth frequency component $M_{k,t}$ switches from the *good* low volatility state to the *bad* high volatility state $(b^*)^{2(\bar{k}-k)}$ times more often under $\mathbb{Q}$ than under $\mathbb{P}$, while a switch from the bad state to the good state occurs $(b^*)^{2(\bar{k}-k)}$ times less often. Consequently, as b^* increases, negative jumps in returns occur more frequently and the expected long-run volatility level increases.

We specify both the equity and volatility risk-premia to be proportional to V_t, as in Pan (2002). Under the risk-neutral measure $\mathbb{Q}$, the log-price process solves the stochastic differential equation:

$$\begin{aligned} ds_t &= (r_t - d_t - V_t/2)\,dt + \sqrt{V_t}dW^*_{1,t} + dJ_t - \bar{J}^*_{M_t}dt, \\ dV_t &= \kappa^*\left(\theta^*_{M_t} - V_t\right)dt + \sigma\sqrt{V_t}\left(\rho dW^*_{1,t} + \sqrt{1-\rho^2}dW^*_{2,t}\right), \end{aligned} \tag{15}$$

where κ^* and $\bar{\theta}^* = \kappa\bar{\theta}/\kappa^*$ are fixed parameters, $(W^*_{1,t}, W^*_{2,t})$ is a Wiener process under the risk-adjusted measure, Q^* is the risk-neutral transition matrix of the Markov chain M_t, and $\theta^*_{M_t} = \bar{\theta}^*\prod_{k=1}^{\bar{k}} M_{k,t}$ is the "long-run" level of stochastic variance. As previously, $N_{i,j,t}$ denotes the number of switches from state i to state j between dates 0 and t. The cumulative jump is $J_t = \sum_{i,j} J_{i,j}N_{i,j,t}$, and the compensator $\bar{J}^*_{M_t}$ is defined by

$$\bar{J}^*_{M_t} = \sum_{i=1}^{d} 1_{\{M_t = m^i\}} \sum_{j=1}^{d} q^*_{i,j}J_{i,j} \tag{16}$$

The jumps of the Skew MSM model are deterministic and therefore remain the same under both measures, but the rate at which jumps occur under $\mathbb{Q}$ is controlled by the rate matrix Q^*. The market price of risk, given by $\alpha\sqrt{V_t}$, is consistent with the CAPM principle of greater risk requiring greater compensation.[10]

In summary, the joint objective and risk-neutral distributions are fully described by eleven parameters, that is by the parameter vector $\Psi_{\mathbb{P}}$ in (12) and by

$$\Psi_{\mathbb{Q}} \equiv (\kappa^*, b^*)\,, \tag{17}$$

where κ^* denotes the risk-neutral adjustment speed of stochastic variance, and b^* controls the risk-adjusted transition rates.

3 Option pricing

Fast option pricing is available when the stock price has a known characteristic function under the risk-neutral measure $\mathbb{Q}$ (e.g., Carr and Madan (1999)). Consider a maturity date $T \geq t$ and let

[10] When the equity risk-premium is given by αdt the market price of risk is $(\alpha - r_t + d_t)/\sqrt{V_t}$ and implies that as volatility approaches its lower bound the market price of risk approaches infinity. This would potentially permit arbitrage opportunities and is the primary motivation for choosing an equity risk-premium given by $\alpha V_t dt$.

$\tau = T - t$ denote time to maturity. The characteristic function of s_T conditional on the state (s_t, V_t, M_t) is given by the Fourier transform:

$$\psi\left(u, \tau; s_t, V_t, M_t\right) = \mathbb{E}^{\mathbb{Q}}\left(e^{\mathrm{i}us_T}\middle| s_t, V_t, M_t\right),$$

where $\mathrm{i} = \sqrt{-1}$, $u \in \mathbb{R}$, and $\tau \in \mathbb{R}_+$. We assume for simplicity that the risk-free rate and the dividend yield are time-invariant: $r_t = r$ and $d_t = d$ for all t.

When the state M_t is constant, Skew MSM coincides with the model of Heston (1993) and the characteristic function is known. Consider the complex-valued functions introduced by Heston (1993):

$$C_i(u,\tau) = (r - d - \bar{J}^*_{m^i})\mathrm{i}u\tau + \frac{\kappa^*\theta^*_{m^i}}{\sigma^2}\left\{\left[\kappa^* - \rho\sigma\mathrm{i}u - \gamma(u)\right]\tau - 2\log\frac{1 - G(u)e^{-\gamma(u)\tau}}{1 - G(u)}\right\}, \quad (18)$$

$$D(u,\tau) = \frac{\kappa^* - \rho\sigma\mathrm{i}u - \gamma(u)}{\sigma^2}\frac{1 - e^{-\tau\,\gamma(u)}}{1 - G(u)e^{-\tau\,\gamma(u)}}, \quad (19)$$

$$\gamma(u) = \sqrt{\left(\kappa^* - \rho\sigma\mathrm{i}u\right)^2 + \sigma^2\left(u^2 + \mathrm{i}u\right)}, \quad (20)$$

$$G(u) = \frac{\kappa^* - \rho\sigma\mathrm{i}u - \gamma(u)}{\kappa^* - \rho\sigma\mathrm{i}u + \gamma(u)}, \quad (21)$$

for every $u \in \mathbb{R}$ and $\tau \in \mathbb{R}_+$. We denote by $\dot{C}_i(u,\tau)$ the time derivative $\partial C_i(u,\tau)/\partial\tau$, which is available in closed-form.[11]

We show in Appendix C that Skew MSM is conditionally affine.

Proposition 2 (Characteristic function of Skew MSM). *The characteristic function of s_T is given by:*

$$\psi\left(u, \tau;\, s, V, M_t = m^i\right) = e^{\mathrm{i}us + D(u,\tau)\,V}\bar{\psi}_i(u,\tau). \quad (23)$$

The d-dimensional function $\bar{\psi}(u,\tau) = [\bar{\psi}_1(u,\tau), \ldots, \bar{\psi}_d(u,\tau)]'$ solves the system of ordinary differential equations:

$$\frac{\partial\bar{\psi}(u,\tau)}{\partial\tau} = \mathbf{\Phi}(u,\tau)\bar{\psi}(u,\tau), \quad (24)$$

where $\mathbf{\Phi}(u,\tau) = \left[\mathbf{\Phi}_{ij}(u,\tau)\right]_{ij}$ is the $d \times d$ matrix with diagonal elements $\mathbf{\Phi}_{ii}(u,\tau) = \dot{C}_i(u,\tau) + q^*_{i,i}$ and off-diagonal elements $\mathbf{\Phi}_{ij}(u,\tau) = q^*_{i,j}\,e^{\mathrm{i}uJ_{ij}}$.

[11] We easily verify that

$$\dot{C}_i\left(u,\tau\right) = (r - d - \bar{J}^*_{m^i})\mathrm{i}u + \frac{\kappa^*\theta^*_{m^i}}{\sigma^2}\left[\kappa^* - \rho\sigma\mathrm{i}u - \gamma(u) - \frac{2\gamma(u)G(u)e^{-\tau\,\gamma(u)}}{1 - G(u)e^{-\tau\,\gamma(u)}}\right], \quad (22)$$

for all u, τ.

The system of ordinary differential equations (24) is linear and first-order with variable coefficients. It can be solved rapidly using standard solvers. A variety of computationally efficient methods have been proposed for pricing options using transforms of the characteristic function. Carr and Madan (1999) develop an equation for the price of European options using a Fast Fourier Transform (FFT) of the characteristic function, and more recently, Fang and Oosterlee (2008) show how the option price can be constructed using a cosine expansion of the characteristic function. The cosine expansion method requires far fewer evaluations of the characteristic function for a given level of accuracy and so we employ it in the rest of the paper.

Skew MSM is conditionally affine for two main reasons: (1) the regime switches affect only the "long run level" θ_{M_t} of the variance process and (2) the functional form of the volatility risk premium ensures that the risk neutral process remains affine. These properties allow us to obtain a tractable option pricing framework. Furthermore, it should be noted that although Skew MSM is conditionally affine, the unconditional process is capable of highly non-linear behavior. Our method therefore represents a tractable method of pricing options with a non-affine process.

4 Estimation methodology

4.1 Euler discretization and filtering equations

We discretize the filtering equations by an Euler scheme which is simple to perform and has been shown to work well in practice.[12] Consider a step size Δt and assume that the stock price is observed at dates $n\,\Delta t$, where $n = 0, 1, ..., \infty$. The discrete-time Markov chain $\{M_n\}$ has a state-space $\mathcal{D} = (m^1, ..., m^d)$ and a transition matrix A with elements $a_{ij} = \mathbb{P}(M_{n+1} = m^j | M_n = m^i)$. The transition matrix for the Markov chain can be found using the rate matrix Q as follows

$$A = \exp\left(Q\,\Delta t\right). \tag{25}$$

By a slight abuse of notation, we henceforth denote by t the integer time index of the discretized process.

[12] See, e.g., Christoffersen, Jacobs and Mimouni (2010).

The discrete-time vector (s_t, v_t) evolves under $\mathbb{P}$ according to the following transition equations:

$$s_t - s_{t-1} = \mu_t \Delta t + \sqrt{v_t \Delta t}\, \varepsilon_{1,t} + \hat{J}\left(M_t, M_{t-1}\right) \tag{26}$$

$$v_{t+1} = v_t + \kappa\left(\theta_{M_{t+1}} - v_t\right)\Delta t + \rho\sigma\sqrt{v_t \Delta t}\varepsilon_{1,t} + \sqrt{1-\rho^2}\sqrt{v_t \Delta t}\, \varepsilon_{2,t}, \tag{27}$$

where $\mu_t = r_t - d_t + (\alpha - 1/2)\, v_t$, and $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are uncorrelated standard Gaussian variables. The discretized jump process is defined as

$$\hat{J}\left(M_t, M_{t-1}\right) = \sum_{i,j} J_{i,j} 1_{\{M_{t-1}=m^i, M_t = m^j\}} - \bar{J}\left(M_t\right),$$

where $J_{i,j}$ is given by (11) and $\bar{J}\left(M_t\right) = \sum_{j=1}^{d} a_{ij}\left(e^{J_{ij}} - 1\right)$ if $M_t = m^i$.

The financial economist observes a measurement vector $y_t \in \mathbb{R}^p$ every period t. We collect this available information at time-step t into the vector $y_{1:t} = (y_1, \ldots, y_t)$. The filtering problem involves recursively calculating the distribution of the latent state (v_t, M_t) conditional on $y_{1:t}$. We introduce two particle filter methods for filtering and estimating Skew MSM. The first is used to estimate the model on equity data only, while the second is useful to estimate the model using both equity and option data. Both filters adapt the sampling and importance resampling filter of Gordon et al. (1993) to a regime-switching jump-diffusion setting.[13]

4.2 Particle filter for stock returns

The observation y_t contains the log-return:

$$y_t = s_t - s_{t-1}.$$

Conditional on (v_t, M_t, M_{t-1}), the observation vector y_t is Gaussian with mean $\mu_t \Delta t + \hat{J}(M_t, M_{t-1})$ and variance $v_t\, \Delta t$. The corresponding density, which is usually called the *observation density*, is denoted by $f(y_t | v_t, M_t, M_{t-1})$.

The observation density helps to define a set of particles $\left(v_t^{(k)}, M_t^{(k)}, M_{t-1}^{(k)}\right)$, $k = 1, \ldots, K$, that targets the joint distribution of (v_t, M_t, M_{t-1}) conditional on $y_{1:t}$. Given the period$-t$ filter, the

[13] Our particle filter methods easily extend to the auxiliary particle filter of Pitt and Shephard (1999). However, we chose to focus on the sampling and importance resampling filter because it has been shown to be effective for equity time-series and is straightforward to implement. We refer the reader to Doucet and Johansen (2011) for a detailed comparison of standard and auxiliary particle filters.

construction for period $t+1$ proceeds in three steps. First, we propagate the particles forward. Second, we observe y_{t+1} and assign to each first-step particle a weight proportional to the observation density. In the final step, we uniformly resample from the approximate posterior distribution in order to ensure that the sample does not disperse over time. A full description of the algorithm is provided in Appendix D.

The particle filter can be computed recursively at dates $t = 1, \ldots, T$. We estimate the likelihood of $y_{1:T}$ by

$$\hat{\mathcal{L}} = \prod_{t=1}^{T} \left[\frac{1}{K} \sum_{k=1}^{K} f(y_t | v_t^{(k)}, M_t^{(k)}, M_{t-1}^{(k)}) \right]. \tag{28}$$

The characteristic function of the stock price s_T conditional on $s_{t'}$, $t' \leq t$, is estimated by

$$\sum_{k=1}^{K} \psi(u, \tau;\, s_t, v_t^{(k)}, M_t^{(k)})/K.$$

We can then compute option prices as explained in Section 3.

4.3 Particle filter for joint filtering of stock and options

We now assume that the financial economist observes every period the log return $s_t - s_{t-1}$ and the log-prices $c_{i,t}$, $i = 1, \ldots, D_t$, of options with different strikes and maturities. The observation vector at date t is therefore:

$$y_t = (s_t - s_{t-1};\, c_{1,t};\, \ldots;\, c_{D_t,t})'.$$

We assume that the log-option price $c_{i,t}$ is observed with Gaussian error, capturing any mispricing due to illiquidity, market frictions, and model misspecification. We write the error structure as

$$c_{i,t} = c_{i,t}^* + \sigma_c \, \eta_{i,t} - \frac{\sigma_c^2}{2}, \tag{29}$$

where $c_{i,t}^*$ is the true unobserved log-option price and $c_{i,t}$ is the market log-option price observed with error. The error term $\eta_{i,t}$ is an i.i.d. standard Gaussian variable and σ_c is the standard deviation of the noise. The correction term $-\sigma_c^2/2$ ensures that in levels, the option price is log-normally distributed about the true option price.

For every (v_t, M_t, M_{t-1}), the observation y_t is conditionally Gaussian and the observation density $f(y_t | v_{t+1}, M_{t+1}, M_t)$ is therefore available in closed-form. When the options data contains a large

number of contracts, however, the evaluation of the observation density is numerically costly. For this reason, we develop a particle filter that reduces the number of evaluations of the observation density. This approach, which is inspired by the stratification method of Kitagawa (1996), is described in Appendix D.

Two-step MLE The joint estimation particle filter requires the computation of $K \times D_t$ model-implied option prices for each likelihood evaluation, where K is the number of particles and D_t is the total number of options in the sample. Although the option pricing method for Skew MSM is relatively fast, it is numerically intractable to perform a full MLE when a large sample of options is used. Hence, rather than performing a full joint estimation over all eleven model parameters, we use a two-step maximum likelihood estimation procedure. The nine objective $\mathbb{P}$-parameters are estimated using a long time-series of returns. Using their ML estimates, we then estimate the remaining two risk-neutral $\mathbb{Q}$-parameters (plus the additional option observation parameter) using the joint data. By Murphy and Topel (1985), the two-step estimation procedure produces an asymptotically consistent estimate of all eleven parameters.

4.4 Benchmark specifications

We compare the in-sample return performance of Skew MSM model with several other models. They fall into three categories: restricted versions of Skew MSM; continuous-time jump-diffusion processes, particularly of the affine variety; and discrete-time GARCH-type models. We briefly specify these models below.

Restricted versions of Skew MSM. We consider a version of the model in which the stochastic variance V_t is non-diffusive and coincides with θ_{M_t}. The corresponding process, which we call MSM with jumps and denote MSM-J, is obtained by setting $\kappa = \infty$ and $\sigma = \rho = 0$. MSM itself is another natural benchmark, which corresponds to also setting $\beta = 0$.

Jump-diffusions. Jump-diffusions are a flexible and widely applied family of continuous time processes. We consider specifications of the form:

$$\frac{dS_t}{S_t} = \mu(V_t)\,dt + \sqrt{V_t}dW_{1,t} + Z_t, \tag{30}$$

$$dV_t = \kappa(\theta - V_t)\,dt + \sigma V_t^b(\rho\,dW_{1,t} + \sqrt{1-\rho^2}\,dW_{2,t}), \tag{31}$$

where $W_{1,t}$ and $W_{2,t}$ are independent Wiener processes, and Z_t is a pure-jump process. We use the following specifications:

SV $Z_t = 0$ and $b = 1/2$.

SVJ0 Z_t is a normal compound Poisson process with arrival rate λ and $b = 1/2$.

SVJ1 Z_t is a normal compound Poisson process with arrival rate $V_t\lambda$ and $b = 1/2$.

NA-SV $Z_t = 0$ and $b = 1$.

The SV, SVJ0, and SVJ1 jump-diffusions are affine, while NA-SV is non-affine. Bates (1996) and Bakshi et al. (1997) provide empirical implementations of SV and SVJ0. Pan (2002) Bates (2006) and Andersen et al. (2012) also consider SVJ1, and the non-affine NA-SV specification has been recently used empirically by Christoffersen, Jacobs and Mimouni (2010).

GARCH benchmarks. GARCH models are stochastic processes where the volatility of returns is a deterministic auto-regressive function of past observations. These models are able to capture some of the volatility persistence observed in asset returns. We consider specifications of the form:

$$s_t - s_{t-1} = \mu_t + \sigma_t\varepsilon_t, \tag{32}$$

where ε_t are *i.i.d.* standard Gaussian and

$$\sigma_t^2 = \omega + \sum_{p=1}^{P}\alpha_p\left|\varepsilon_{t-p}\right|^\delta + \sum_{o=1}^{O}\gamma_o\left|\varepsilon_{t-o}\right|^\delta \mathbf{1}_{\{\varepsilon_{t-o}<0\}} + \sum_{q=1}^{Q}\beta_q\sigma_{t-q}^\delta. \tag{33}$$

The integers P, O, and Q determine the number of model parameters. The variance specification in (33) nests different GARCH-type models and builds the basis for our empirical investigation. In particular, we explore the following cases:

GARCH $P = 1, O = 0, Q = 1, \delta = 2$

ZARCH $P = 1, O = 1, Q = 1, \delta = 1$

GJR-GARCH $P = 1, O = 1, Q = 1, \delta = 2$

The classic GARCH(1,1) model is symmetric. ZARCH (sometimes called TGARCH) allows the conditional standard deviation to depend upon the sign of the lagged innovations ε_t (Zakoian (1994)). Similarly, the GJR-GARCH model of Glosten et al. (1993) assumes that the effect of ε_t^2 on the conditional variance σ_t^2 depends on the sign of ε_t. GJR-GARCH and ZARCH introduce these dependencies to accommodate the leverage effect.

5 Empirical results

We begin by reporting empirical results based on stock return data only. We then include option data into the estimation and conduct the in- and out-of-sample evaluation of Skew MSM.

5.1 Estimation on equity data

We estimate the models using a sample of daily S&P 500 log-returns from March 6, 1957 to September 28, 2007. We convert the returns into log excess returns using monthly Treasury bill returns from Ibbotson and Associates, downloaded from Ken French's website.

Table 1 reports ML estimation results for Skew MSM. The rows of the table correspond to the number of frequency components $\bar{k}$. Hence, the first row corresponds to the one component model, in which the long-run volatility switches between two values, and there is a single deterministic jump value associated with the switch. The estimated value for $\lambda_{\bar{k}}$ of 0.007 corresponds to a switch in long-run volatility occurring approximately on average 1.7 times every year. The return jump associated this switch is moderately large with a magnitude of 0.035. The value for m_0 of 1.04 is near to 1 meaning that the two long-run volatility levels are quite close together. Most of the volatility variation is captured by the diffusive volatility component and the regime-switching is used primarily to capture return outliers.

Table 1: Maximum likelihood estimation of Skew MSM

$\bar{k}$	α	b	m_0	$\lambda_{\bar{k}}$	$\bar{\theta}$	κ	σ	ρ	β	ln L
1	1.01	-	1.04	6.85E-3	0.020	3.0	0.26	-0.56	-2.38E-4	44006
2	1.00	5.87	1.16	8.81E-3	0.020	3.8	0.25	-0.57	-3.05E-5	44047
3	1.15	5.50	1.27	2.00E-2	0.022	5.0	0.29	-0.58	-1.30E-4	44057
4	1.00	3.11	1.40	3.78E-2	0.024	25.6	0.23	-0.86	-1.42E-4	44073
5	2.15	5.00	1.41	2.22E-1	0.022	19.7	0.17	-0.98	-7.92E-5	44089
6	1.01	1.64	1.34	3.25E-2	0.030	34.7	0.20	-0.99	-1.03E-4	44098
7	1.14	1.96	1.33	3.29E-2	0.022	42.6	0.20	-0.99	-1.19E-4	44108

The table reports the ML estimation results for the Skew MSM model using daily S&P 500 log excess returns from March 6, 1957 to September 28, 2007. Rows correspond to the number of frequency components of the estimated model. The likelihood increases monotonically in the number of frequencies.

As additional frequency components are added, several trends emerge. First, b decreases and $\lambda_{\bar{k}}$ increases, indicating that changes occur at higher frequency and the spacing of components across frequencies becomes closer. These findings are consistent with empirical results for standard MSM processes (e.g., Calvet and Fisher (2004).) As more components are added, m_0 increases until stabilizing and then slightly declining for large $\bar{k}$. The adjustment speed κ of the diffusive volatility component strongly increases with $\bar{k}$, indicating that the non-diffusive θ_{M_t} increasingly dominates movements in volatility.

A more detailed look at the volatility diffusion parameters κ, σ, and ρ shows interesting patterns. When $\bar{k}$ is low, the adjustment speed κ of the diffusive volatility component indicates a half-life on the order of months, and the correlation ρ between volatility and returns is -0.56, similar to values obtained for standard stochastic volatility models. However, as the number of frequency components increases, the volatility starts to mean revert more strongly around the MSM process. For large values of $\bar{k}$, the adjustment speed κ is much higher and the diffusive volatility shocks decay in about a week. Moreover, for large k the volatility diffusion is almost perfectly negatively correlated with the return innovations. For $\bar{k} \geq 5$, the conditional volatility for horizons of one month and longer is almost entirely captured by the MSM components, since at these horizons the diffusive volatility component integrates out. The role played by the diffusive volatility component is then to capture transitory shocks to volatility, which are highly correlated with shocks to returns.

The estimated jump parameter value β ensures that the largest jump component is roughly the

same (approximately 0.2 in magnitude) no matter how many frequency components are added. As more frequency components are added, smaller but more frequent jumps are included into Skew MSM. In addition to capturing the large return outliers, the jumps introduce dependency between the returns and the long-run volatility level. This particular feature is difficult to replicate outside the regime-switching framework.

Finally, we examine the behavior of the log-likelihood. As $\bar{k}$ increases from one to seven, the number of parameters remains the same, but the likelihood increases monotonically, consistent with standard MSM. The likelihood function increases rapidly at low values of $\bar{k}$ and increase more slowly for low values of $\bar{k}$, consistent with the convergence result in Proposition 1. The increases in likelihood is nonetheless substantial even for $\bar{k} = 7$, the largest value we consider.

We estimate the alternative models and report the results in Table **??** of Appendix E. All estimates are comparable to the values obtained in the literature. Table 2 compares the likelihood of each alternative models with Skew MSM. The Skew MSM likelihood is uniformly larger than the alternatives, but also has at least one more parameter than each alternative model. To adjust for this, we apply the Schwartz-Bayesian information criterion (BIC). Under this criterion, Skew MSM is preferred to all alternatives, with the NA-SV, MSM-J, SVJ1, and SVJ0 being the next best models.

We also test the significance of the BIC differences using the test of Vuong (1989). The null hypothesis is that each of the alternative models fits the data equally well as Skew MSM. Column 5 of Table 2 reports p-values for the Vuong tests. All of the alternative models are rejected in favor of Skew MSM with high significance. We also report p-values for a HAC-adjusted variation of the Vuong test and find that the conclusions remain unchanged.

Finally, we investigate the ability to match autocorrelation and cross-correlation functions. For simplicity, we restrict this analysis to Skew MSM and SVJ1. We first simulate returns for these two models using the estimated parameter values in Table 2. In Figure 2 we plot the resulting autocorrelation and cross correlation functions of the simulated data, along with the same functions for the empirical data. Skew MSM captures both the exponentially decaying cross-correlation and hyperbolically decaying variance autocorrelation found in the data. In contrast, SVJ1 has a variance autocorrelation function that decays too quickly and a cross-correlation function that decays to

Table 2: In-sample model comparison

	No. of parameters	ln L	BIC	BIC p-value vs. Skew MSM	
				Vuong (1989)	HAC Adj
Skew MSM(7)	9	44108	-88131		
MSM-J(7)*	6	44018	-87980	6.31E-7	2.99E-3
MSM(7)*	5	43960	-87873	1.58E-6	1.76E-3
SV	5	43970	-87893	2.23E-2	1.16E-1
NA-SV	5	44041	-88036	9.65E-2	2.14E-1
SVJ0	8	44047	-88018	2.69E-3	6.00E-2
SVJ1	8	44064	-88052	2.05E-2	1.18E-1
GARCH*	4	43683	-87329	5.88E-7	2.04E-3
GJR-GARCH*	5	43803	-87558	8.39E-5	1.43E-2
ZARCH*	5	43831	-87615	6.22E-4	2.82E-2

The table summarizes information about in-sample goodness-of-fit for the different models. The Bayesian information criterion is given by $BIC = -2 \ln L + NP \ln T$. The analysis is based on S&P 500 daily returns over the in-sample period (March 6, 1957-September 28, 2007) and contains 12,730 observations. The last two columns give the p-values from a test that the corresponding model dominates the Skew MSM model by the BIC. The first value uses the Vuong (1989) methodology and the second adjusts the test for heteroscedasticity and autocorrelation. A low p-value indicates that the corresponding model would be rejected in favor of the Skew MSM model. The models marked by * are estimated using a closed form likelihood function. The other models are estimated using a particle filter.

slowly. Hence, the good performance of Skew MSM can be related to important statistical properties of equity data.

5.2 Empirical results using return and option data

Option prices contain a rich set of information about the conditional distribution of the underlying price process, which should be useful for filtering and identification of latent states. Following this intuition, we estimate the leading latent state models in our set of alternatives, Skew MSM, SVJ0, and SVJ1, on both returns and options. To the best of our knowledge, we are the first to carry out joint filtering and ML estimation using both options and returns for any of these models, including the benchmark models.

We use S&P 500 index call option monthly data from the January 1998 to September 2010 period for our option-based empirical analysis. Hence, we include the period of the recent financial crisis. Estimating the model jointly on option data is computationally challenging. We select option

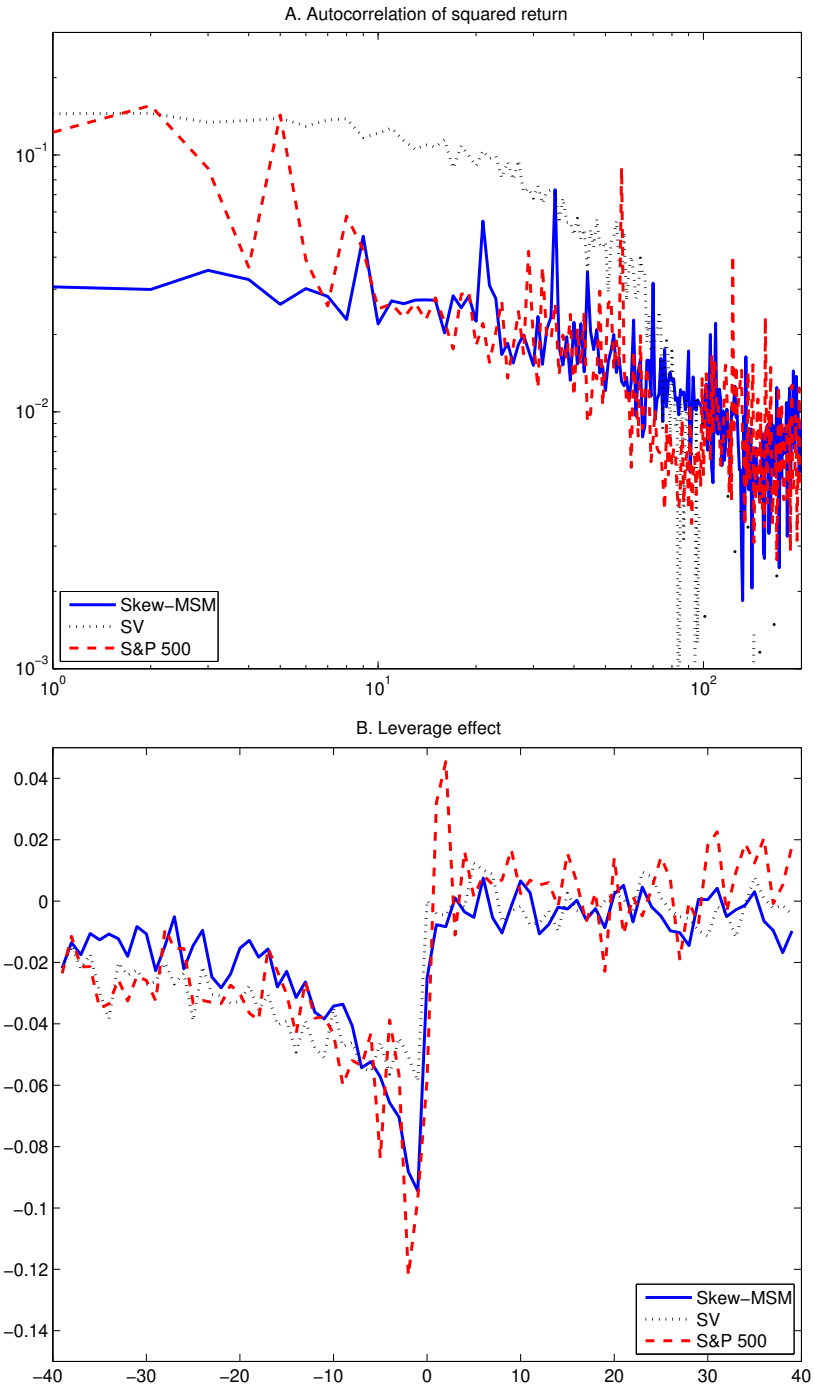

Figure 2: Auto- and cross-correlation plots. The plots show moments of the Skew MSM and SV models using simulated data, and compares them to S&P 500 returns. Panel A is a log-log plot of the autocorrelation function of squared returns. Panel B shows the cross-correlation between returns and squared returns.

contracts from the first Wednesday of each month in the sample with maturities between 0.2 and 1 year and with a moneyness levels between 0.9 and 1.1. We thus obtain a series of 117 option surfaces.

Option surfaces tend to be highly persistent, and so the large gap between each option surface in our sample ensures that we get a wide variety of implied volatility surfaces with which to estimate the risk-premium parameters.

Panel A of Table 3 reports summary statistics for the full set of options in our in-sample period (January 1998–September 2007). There are 5028 contracts in this data set. The average implied volatility values show the familiar volatility smirk, although it is less pronounced at longer maturities. Each of the option surfaces in the sample contains between 30 and 60 option contracts. Panel B gives summary statistics for the out-of-sample period (October 2007 – September 2010). The options again have maturities between 0.2 and 1 year and moneyness levels between 0.9 and 1.1, and the sample contains 2,915 contracts.

Table 4 reports the second stage ML parameter estimates in our in-sample data spanning 1998-2007 for SVJ0, SVJ1, and for Skew MSM with $\bar{k} \in \{1, 2, 3\}$. For both SVJ models, we find modest jump-risk premia, consistent with the Pan (2002) and Christoffersen, Jacobs and Mimouni (2010). Under the risk-neutral measure, the mean jump frequency is 0.62 compared to 0.45 under the objective measure. The volatility risk-premium in these models is quite large, again consistent with Christoffersen, Jacobs and Mimouni (2010). The Skew MSM risk premium implies that the probability of the low frequency component switching from a low value to a high value is 1.4 times as high under the risk-neutral measure.

The option observation error parameter estimate σ_c is considerably lower for Skew MSM than for the SVJ models, indicating that Skew MSM provides a better in-sample fit. The log-likelihoods and BIC values confirm these results. Skew MSM(3) model has the highest likelihood and by far the lowest BIC, and the Vuong tests show that the difference in BIC values is highly significant.

Table 5 shows RMSE for Skew MSM(3) and the jump-diffusion models in the in-sample option data. Skew MSM has a lower RMSE than the other models across almost all maturities and moneyness levels, with particularly strong outperformance for out-of-the money options. The total RMSE of 0.065 for Skew MSM is an improvement of about 33% over the SVJ0 model. For comparison, Christoffersen, Jacobs and Mimouni (2010) find an improvement relative to the same benchmark of only about 10% for a non-affine stochastic volatility model. Hence, Skew MSM dominates the

Table 3: S&P 500 index call option data summary

A. In-sample period (Jan. 1998-Sept. 2007)

		Out of the money	2	3	4	5	In the money	All strikes
$0.2<\tau<0.4$	No. of options	510	277	317	284	259	421	2068
	Av. price ($)	14.92	28.16	43.88	62.47	83.54	114.02	56.43
	Av. BSIV	0.169	0.176	0.186	0.196	0.203	0.221	0.191
$0.4<\tau<0.6$	No. of options	253	115	135	117	112	191	923
	Av. price ($)	29.29	46.43	65.12	84.39	102.63	132.68	73.95
	Av. BSIV	0.168	0.175	0.183	0.196	0.197	0.209	0.187
$0.6<\tau<1$	No. of options	553	247	299	253	245	440	2037
	Av. price ($)	50.11	68.19	88.31	106.18	125.09	154.37	96.41
	Av. BSIV	0.177	0.183	0.189	0.197	0.197	0.211	0.191
All maturities	No. of options	1316	639	751	654	616	1052	5028
	Av. price ($)	32.47	46.92	65.39	83.30	103.54	134.28	75.84
	Av. BSIV	0.17	0.18	0.19	0.20	0.20	0.21	0.19

B. Out-of-sample period (Oct. 2007-Sept. 2010)

		Out of the money	2	3	4	5	In the money	All strikes
$0.2<\tau<0.4$	No. of options	400	203	199	188	170	290	1450
	Av. price ($)	18.78	35.54	49.15	68.88	82.61	111.65	57.85
	Av. BSIV	0.215	0.234	0.240	0.257	0.260	0.279	0.245
$0.4<\tau<0.6$	No. of options	133	75	76	66	66	112	528
	Av. price ($)	34.58	55.29	71.66	88.57	101.73	127.21	77.65
	Av. BSIV	0.219	0.249	0.250	0.266	0.267	0.282	0.253
$0.6<\tau<1$	No. of options	266	122	120	109	94	226	937
	Av. price ($)	54.15	76.32	93.87	108.88	121.48	149.96	98.35
	Av. BSIV	0.226	0.246	0.245	0.266	0.257	0.269	0.249
All maturities	No. of options	799	400	395	363	330	628	2915
	Av. price ($)	33.19	51.68	67.06	84.47	97.51	128.21	74.45
	Av. BSIV	0.22	0.24	0.24	0.26	0.26	0.28	0.248

Panel A shows summary statistics for the option data of our in-sample period. We select a set of 117 option surfaces, defined as the first Wednesday of each month between January 1998 and September 2007. We categorize the data into different moneyness buckets defined as (1) : $0.9 \leq S/K \leq 0.95$, (2) : $0.95 < S/K \leq 0.975$, (3) : $0.975 < S/K \leq 1.0$, (4) : $1.0 < S/K \leq 1.025$, (5) : $1.025 < S/K \leq 1.05$, (6) : $1.05 < S/K \leq 1.1$, and different maturity buckets. Panel B shows summary statistics for the out-of-sample period, defined as the first Wednesday of each month between October 2007 and September 2010.

Table 4: In-sample maximum likelihood estimation on option data

	κ^*	b^*	σ_c	ln L	BIC	BIC p-value vs. Skew MSM Vuong (1989)	Skew MSM HAC adj
Skew MSM(3)	2.70	1.24	0.087	5978	-11899		
Skew MSM(2)	1.62	1.21	0.090	5354	-10651	6.22E-6	2.81E-4
Skew MSM(1)	1.38	1.35	0.094	4912	-9767	3.18E-7	7.49E-4
	κ^*	λ^*	σ_c	ln L			
SV	2.77		0.132	3565	-7091	6.11E-13	6.61E-4
SVJ0	0.70	0.62	0.093	4986	-9920	1.30E-11	4.81E-3
SVJ1	1.27	65.1	0.095	4872	-9692	1.12E-11	2.51E-3

The table reports the risk-premia and option observation error parameters for Skew MSM with 1 to 3 frequency components, SVJ0, and SVJ1 estimated using a two-step ML method on a panel of monthly option data over the in-sample period (January 1998-September 2007). The Bayesian information criterion is given by $BIC = -2 \ln L + NP \ln T$. The last two columns give the p-values from a test that the corresponding model dominates the Skew MSM model by the BIC. The first value uses the Vuong (1989) methodology and the second adjusts the test for heteroscedasticity and autocorrelation. A low p-value indicates that the corresponding model would be rejected in favor of Skew MSM.

benchmark SVJ0 model by the widest margin in the existing literature.

Table 6 shows RMSE in the out-of-sample data. Again, Skew MSM considerably outperforms both SVJ0 and SVJ1, with an RMSE about 23% lower than the next best SVJ0 model. These results confirm that the good performance of Skew MSM is not due to in-sample overfitting. The model is stable and performs well in out-of-sample tests.

To better understand the good performance of Skew MSM, Figure 3 shows the implied volatility smiles for Skew MSM(3) and SVJ1 with the parameters given in Table 2. For the plot, we fix the spot volatility level. In Panel A, we plot the implied volatility smiles for Skew MSM(3) in the maximum volatility state $M_t = (m_0, m_0, m_0)$, in which all volatility components are high. In Panel B and C, we plot the corresponding graphs for the intermediate volatility state $M_t = (2 - m_0, m_0, m_0)$ and the minimum volatility state $M_t = (2 - m_0, 2 - m_0, 2 - m_0)$, respectively. In Panel D, we plot the implied volatility smiles for SVJ1. Holding spot volatility constant, depending on the state of M_t we can generate different shapes of implied volatility smiles at different maturities. In contrast, for SVJ1 once we fixed the spot volatility we can only generate one set of implied volatility shapes. Note that, if we were to increase the state space of M_t, we could further enrich the set of possible implied volatility shapes at a given spot volatility. We leave this extension open for future research.

Table 5: In-sample call option log-price RMSE by moneyness and maturity

		Out of the money	2	3	4	5	In the money	All strikes
$0.2<\tau<0.4$	Skew MSM(3)	0.068	0.068	0.066	0.065	0.060	0.050	0.063
	SV	0.148	0.080	0.080	0.071	0.064	0.050	0.094
	SVJ0	0.111	0.099	0.095	0.082	0.075	0.059	0.090
	SVJ1	0.108	0.090	0.091	0.080	0.074	0.057	0.086
$0.4<\tau<0.6$	Skew MSM(3)	0.064	0.059	0.060	0.059	0.058	0.050	0.059
	SV	0.151	0.097	0.086	0.078	0.075	0.061	0.104
	SVJ0	0.090	0.084	0.079	0.073	0.076	0.060	0.078
	SVJ1	0.099	0.085	0.081	0.074	0.075	0.059	0.082
$0.6<\tau<1$	Skew MSM(3)	0.094	0.059	0.060	0.059	0.054	0.053	0.069
	SV	0.180	0.113	0.101	0.094	0.079	0.072	0.112
	SVJ0	0.126	0.072	0.067	0.067	0.061	0.058	0.086
	SVJ1	0.137	0.080	0.074	0.072	0.063	0.059	0.092
All maturities	Skew MSM(3)	0.080	0.063	0.063	0.062	0.058	0.051	0.065
	SV	0.163	0.097	0.090	0.082	0.073	0.062	0.108
	SVJ0	0.114	0.087	0.082	0.075	0.070	0.059	0.086
	SVJ1	0.119	0.085	0.082	0.076	0.070	0.059	0.088

We report the RMSE of call option prices over the in-sample period (Jan. 1998-Sept 2007). The RMSE is calculated from the difference between the model and market log call prices. All parameters are estimated in sample on the S&P 500 index return and option data. We categorize the data into different moneyness buckets defined as (1) : $0.9 \leq S/K \leq 0.95$, (2) : $0.95 < S/K \leq 0.975$, (3) : $0.975 < S/K \leq 1.0$, (4) : $1.0 < S/K \leq 1.025$, (5) : $1.025 < S/K \leq 1.05$, (6) : $1.05 < S/K \leq 1.1$, and different maturity buckets.

Table 6: Out-of-sample call option log-price RMSE by moneyness and maturity

		Out of the money	2	3	4	5	In the money	All strikes
$0.2<\tau<0.4$	Skew MSM(3)	0.099	0.054	0.072	0.079	0.076	0.069	0.079
	SV	0.179	0.071	0.069	0.072	0.068	0.062	0.111
	SVJ0	0.139	0.061	0.082	0.090	0.086	0.078	0.099
	SVJ1	0.144	0.060	0.076	0.084	0.081	0.073	0.099
$0.4<\tau<0.6$	Skew MSM(3)	0.049	0.057	0.081	0.078	0.086	0.078	0.071
	SV	0.107	0.103	0.120	0.107	0.116	0.099	0.108
	SVJ0	0.061	0.075	0.095	0.091	0.102	0.091	0.085
	SVJ1	0.069	0.077	0.099	0.093	0.102	0.090	0.087
$0.6<\tau\leq1$	Skew MSM(3)	0.067	0.081	0.079	0.089	0.089	0.086	0.080
	SV	0.176	0.174	0.156	0.158	0.153	0.133	0.159
	SVJ0	0.084	0.101	0.097	0.104	0.109	0.099	0.097
	SVJ1	0.104	0.114	0.108	0.115	0.116	0.104	0.108
All maturities	Skew MSM(3)	0.082	0.064	0.076	0.082	0.082	0.077	0.078
	SV	0.168	0.118	0.112	0.111	0.109	0.099	0.128
	SVJ0	0.112	0.078	0.089	0.095	0.097	0.088	0.096
	SVJ1	0.122	0.083	0.091	0.096	0.096	0.088	0.100

We report the RMSE of call option prices in the out-of-sample period (Oct. 2007-Sept. 2010). The RMSE is calculated from the difference between the model and market log call prices. All parameters are estimated in sample on the S&P 500 index return and option data. We categorize the data into different moneyness buckets defined as (1) : $0.9 \leq S/K \leq 0.95$, (2) : $0.95 < S/K \leq 0.975$, (3) : $0.975 < S/K \leq 1.0$, (4) : $1.0 < S/K \leq 1.025$, (5) : $1.025 < S/K \leq 1.05$, (6) : $1.05 < S/K \leq 1.1$, and different maturity buckets.

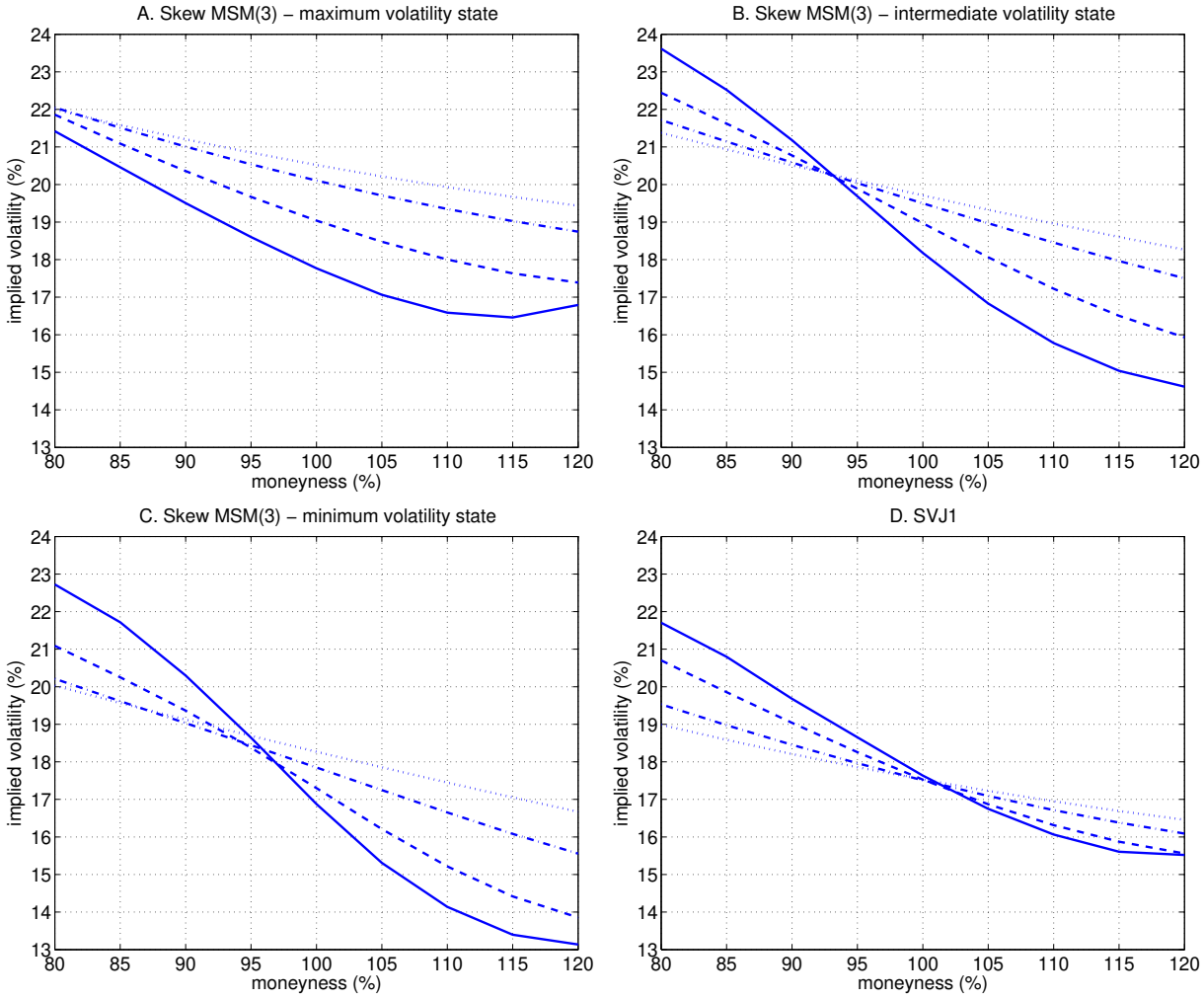

Figure 3: Implied volatility surfaces. The plots show the implied volatility surfaces for the Skew MSM(3) model (panels A to C) and SVJ1 (panel D). For all plots, we fix the spot volatility. Panel A assumes that Skew MSM is in the maximum volatility state $M_t = (m_0, m_0, m_0)$, panel B that it is in the intermediate volatility state $M_t = (2-m_0, m_0, m_0)$, and panel C that is in the minimum volatility state $M_t = (2-m_0, 2-m_0, 2-m_0)$. To calculate the implied volatility, we use the estimated parameter values in Table 2. We plot the implied volatility curves for maturities of three months (solid line), six months (dashed line), one year (dash-dotted line), and 18 months (dotted line).

6 Conclusion

In this paper, we have developed a parsimonious model of equity returns in which shocks of heterogeneous frequency drive the long-run volatility of a standard stochastic volatility model. Shifts in long-run volatility also generate jumps in the return process producing a volatility feedback effect. The model extends the Markov-switching multifractal process of Calvet and Fisher (2004), retaining

its key properties of heterogeneous volatility persistence, multiscaling, and hyperbolically decaying variance autocorrelation, while also incorporating the "leverage" effect and dependence between volatility states and price jumps.

We have estimated the model on both return and option data using a particle filter. In-sample results using return data indicate that the likelihood function increases with the number of volatility components. With a large number of volatility states, the diffusive volatility component mean-reverts strongly to the base MSM process, and is almost perfectly negatively correlated with innovations to returns.

We have compared Skew MSM to affine jump-diffusions and asymmetric GARCH-type models, which are chosen because of previously demonstrated good performance on equity data. The in-sample likelihood is significantly higher for Skew MSM than all other models. We extend our estimation to option data to identify risk-premium parameters and to better filter the latent state variables. The joint likelihood of Skew MSM is significantly higher than the affine jump diffusion models, with a RMSE 30% lower than a standard benchmark, substantially improving on prior literature. The outperformance also extends to an out-of-sample analysis. Skew MSM therefore offers a promising new approach to the joint modeling of equity returns and options.

A Proof of Proposition 1

We sequentially show the convergence of the variance process and the stock price in $L^2(\Omega \times [0,T])$ as the number of components $\bar{k}$ goes to infinity. To facilitate the analysis, we make explicit the dependence of all processes with respect to $\bar{k}$, and denote by $Z_t = \rho W_{1,t} + \sqrt{1-\rho^2} W_{2,t}$ the univariate Wiener process that drives the variance $V_{\bar{k},t}$.

A.1 Convergence of the variance process

Ito's lemma implies that $d(e^{\kappa t} V_{\bar{k},t}) = e^{\kappa t} dV_{\bar{k},t} + \kappa e^{\kappa t} V_{\bar{k},t} dt = e^{\kappa t} \kappa \theta_{M_{\bar{k},t}} dt + \sigma e^{\kappa t} \sqrt{V_{\bar{k},t}} dZ_t$. The variance process therefore satisfies:

$$V_{\bar{k},t} = e^{-\kappa t} V_0 + \kappa \int_0^t e^{-\kappa(t-s)} \theta_{M_{\bar{k},s}} ds + \sigma \int_0^t e^{-\kappa(t-s)} \sqrt{V_{\bar{k},s}} dZ_s. \tag{34}$$

The unconditional expectation of $V_{\bar{k},t}$,

$$\mathbb{E}\left(V_{\bar{k},t}\right) = e^{-\kappa t} V_0 + \kappa \bar{\theta} \int_0^t e^{-\kappa(t-s)}\, ds = e^{-\kappa t} V_0 + \kappa\, \bar{\theta} (1 - e^{-\kappa\, t}),$$

is independent of $\bar{k}$, an invariance property that plays a key role in the rest of the proof. The unconditional variance is

$$\mathrm{Var}\left(V_{\bar{k},t}\right) = A_t + B_{\bar{k},t}.$$

where $A_t = \sigma^2 \int_0^t e^{-2\,\kappa\,(t-s)}\, \mathbb{E}(V_{\bar{k},s})\, ds$ and

$$B_{\bar{k},t} = \kappa^2 \int_0^t \int_0^t e^{-\kappa(t-s)} e^{-\kappa(t-u)} \mathbb{E}[(\theta_{M_{\bar{k},s}} - \bar{\theta})(\theta_{M_{\bar{k},u}} - \bar{\theta})] ds\, du.$$

We note that

$$\mathbb{E}[(\theta_{M_{\bar{k},s}} - \bar{\theta})(\theta_{M_{\bar{k},u}} - \bar{\theta})] = \bar{\theta}^2 \prod_{k=1}^{\bar{k}} \mathbb{E}\left(M_{\bar{k},s}\, M_{\bar{k},u}\right) - \bar{\theta}^2 = \bar{\theta}^2 \prod_{k=1}^{\bar{k}} \left[1 + e^{-\lambda_k\, |s-u|} \mathrm{Var}(M)\right] - \bar{\theta}^2.$$

Hence

$$B_{\bar{k},t} = \kappa^2\, \bar{\theta}^2 \int_0^t \int_0^t e^{-\kappa(t-s)} e^{-\kappa(t-u)} \left\{ \prod_{k=1}^{\bar{k}} \left[1 + e^{-\lambda_k\, |s-u|} \mathrm{Var}(M)\right] - 1 \right\} ds\, du,$$

We observe that $B_{\bar{k},t}$ and $\mathrm{Var}\left(V_{\bar{k},t}\right) = A_t + B_{\bar{k},t}$ both increase with $\bar{k}$.

We now consider

$$\mathbb{E}\left[\left(V_{\bar{k}+1,t} - V_{\bar{k},t}\right)^2\right] = \mathrm{Var}\left(V_{\bar{k}+1,t} - V_{\bar{k},t}\right).$$

We note that $\mathrm{Var}(V_{\bar{k}+1,t}) = \mathrm{Var}(V_{\bar{k}+1,t} - V_{\bar{k},t}) + \mathrm{Var}(V_{\bar{k},t}) + 2\mathrm{Cov}(V_{\bar{k}+1,t} - V_{\bar{k},t}; V_{\bar{k},t})$ and therefore

$$\mathbb{E}\left[\left(V_{\bar{k}+1,t} - V_{\bar{k},t}\right)^2\right] = B_{\bar{k}+1,t} - B_{\bar{k},t} + 2\mathbb{E}[V_{\bar{k},t}(V_{\bar{k},t} - V_{\bar{k}+1,t})]. \tag{35}$$

We now examine separately $B_{\bar{k}+1,t} - B_{\bar{k},t}$ and $2\mathbb{E}[V_{\bar{k},t}(V_{\bar{k},t} - V_{\bar{k}+1,t})]$.

Lemma 1. *The inequalities*

$$0 \leq B_{\bar{k}+1,t} - B_{\bar{k},t} \leq 2t \left[\frac{\mathbb{E}(M^2)}{b}\right]^{\bar{k}} \; Var(M)\kappa^2\,\bar{\theta}^2$$

hold for every $\bar{k}$ and t.

Proof. We note that

$$B_{\bar{k}+1,t} - B_{\bar{k},t} = \kappa^2\,\bar{\theta}^2 \int_0^t \int_0^t e^{-\kappa(t-s)} e^{-\kappa(t-u)}\, e^{-\lambda_{\bar{k}+1}|s-u|}\mathrm{Var}(M) \prod_{k=1}^{\bar{k}} \left[1 + e^{-\lambda_k|s-u|}\mathrm{Var}(M)\right] ds\, du,$$

and therefore

$$B_{\bar{k}+1,t} - B_{\bar{k},t} \leq [\mathbb{E}(M^2)]^{\bar{k}}\,\mathrm{Var}(M)\kappa^2\,\bar{\theta}^2 \int_0^t \int_0^t e^{-\lambda_{\bar{k}+1}|s-u|}\, ds\, du,$$

We verify that

$$\int_0^t \int_0^t e^{-\lambda_{\bar{k}+1}|s-u|}\, ds\, du = \frac{2}{\lambda_{\bar{k}+1}^2}(e^{-\lambda_{\bar{k}+1}t} - 1 + \lambda_{\bar{k}+1}t) \leq \frac{2t}{\lambda_{\bar{k}+1}} = \frac{2t}{\lambda_1 b^{\bar{k}}}$$

and conclude that the lemma holds. ■

In order to construct an upper bound for $2\mathbb{E}[V_{\bar{k},t}(V_{\bar{k},t} - V_{\bar{k}+1,t})]$, it is useful to derive the following result.

Lemma 2. *Let X_t denote the diffusion:*

$$dX_t = \sigma^2\, e^{\kappa t}\, X_t^2 dt - \sigma\, e^{\kappa t/2}\, X_t^{3/2}\, dZ_t.$$

with initial condition $X_0 = 1/V_0$. The inequality

$$\frac{1}{V_{\bar{k},t}} \leq e^{\kappa t}\, X_t \tag{36}$$

holds almost surely. Furthermore,

$$\mathrm{Cov}[X_t, (V_{\bar{k}+1,t} - V_{\bar{k},t})^2] \leq 0 \tag{37}$$

for all t and $\bar{k}$.

Proof. By Ito's lemma, the process $X_{\bar{k},t} = e^{-\kappa t}/V_{\bar{k},t}$ satisfies the stochastic differential equation:

$$dX_{\bar{k},t} = (\sigma^2 - \kappa\theta_{M_{\bar{k},t}})e^{\kappa t}\, X_{\bar{k},t}^2 dt - \sigma\, e^{\kappa t/2}\, X_{\bar{k},t}^{3/2} dZ_t.$$

The processes $X_{\bar{k},t}$ and X_t have identical volatility functions, while the drift function of $X_{\bar{k},t}$ is bounded above by the drift function of X (that is, $(\sigma^2 - \kappa\theta_{M_{\bar{k},t}})e^{\kappa t}\, x^2 \leq \sigma^2 e^{\kappa t}\, x^2$ for all realizations of x, t, and $M_{\bar{k},t}$). We infer from Theorem 1 in Ikeda and Watanabe (1977) that inequality (36) holds almost surely.

Furthermore, Ito's lemma implies that

$$X_t^{-1} = V_0 + \sigma \int_0^t \frac{e^{\kappa s/2}}{\sqrt{V_s}} dZ_s$$

and

$$\begin{aligned}(V_{\bar{k}+1,t} - V_{\bar{k},t})^2 \;\; &= \;\; \int_0^t e^{-2\kappa(t-s)} \left[2\kappa\theta_{M_{\bar{k},s}}(V_{\bar{k}+1,s} - V_{\bar{k},s})(M_{\bar{k}+1,s} - 1) + \sigma^2(\sqrt{V_{\bar{k}+1}} - \sqrt{V}_{\bar{k},s})^2\right] ds \\ &\quad + 2\sigma \int_0^t e^{-2\kappa(t-s)}\, (V_{\bar{k}+1,s} - V_{\bar{k},s})(\sqrt{V}_{\bar{k}+1,s} - \sqrt{V}_{\bar{k},s})\, dZ_s.\end{aligned}$$

Hence $\mathrm{Cov}[X_t^{-1}, (V_{\bar{k}+1,t} - V_{\bar{k},t})^2] \geq 0$ and we conclude that the lemma holds. ■

We can then show

Lemma 3. *The inequality*

$$2\,\mathbb{E}[V_{\bar{k},t}(V_{\bar{k},t} - V_{\bar{k}+1,t})] \leq \sigma^2 e^{-\kappa t} \int_0^t e^{-\kappa(t-s)} \mathbb{E}\,(X_s)\; \mathbb{E}\left[(V_{\bar{k}+1,s} - V_{\bar{k},s})^2\right]\, ds \tag{38}$$

holds for all k and t.

Proof. The integral representation (34) implies that

$$V_{\bar{k},t} - V_{\bar{k}+1,t} = \kappa \int_0^t e^{-\kappa(t-s)} \theta_{M_{\bar{k},s}}(1 - M_{\bar{k}+1,s})\, ds + \sigma \int_0^t e^{-\kappa(t-s)}\, (\sqrt{V_{\bar{k},s}} - \sqrt{V_{\bar{k}+1,s}}) dZ_s.$$

Since $V_{\bar{k},t}$ and $M_{\bar{k}+1,s}$ are independent, we infer that

$$2\mathbb{E}\left[V_{\bar{k},t}(V_{\bar{k},t} - V_{\bar{k}+1,t})\right] = 2\sigma\mathbb{E}\left[V_{\bar{k},t} \int_0^t e^{-\kappa(t-s)}\, (\sqrt{V_{\bar{k},s}} - \sqrt{V_{\bar{k}+1,s}})\, dZ_s\right],$$

which can be rewritten as

$$2\mathbb{E}\left[V_{\bar{k},t}(V_{\bar{k},t}-V_{\bar{k}+1,t})\right]=2\sigma^2\int_0^t e^{-2\kappa(t-s)}\,\mathbb{E}\left[\sqrt{V_{\bar{k},s}}(\sqrt{V_{\bar{k},s}}-\sqrt{V_{\bar{k}+1,s}})\right]ds.$$

Since $\mathbb{E}(V_{\bar{k},s})=\mathbb{E}(V_{\bar{k}+1,s})$, we infer that

$$2\,\mathbb{E}[\sqrt{V_{\bar{k},s}}(\sqrt{V_{\bar{k},s}}-\sqrt{V_{\bar{k}+1,s}})]=\mathbb{E}[(\sqrt{V_{\bar{k},s}}-\sqrt{V_{\bar{k}+1,s}})^2]=\mathbb{E}\left[\frac{(V_{\bar{k},s}-V_{\bar{k}+1,s})^2}{(\sqrt{V_{\bar{k},s}}+\sqrt{V_{\bar{k}+1,s}})^2}\right]. \tag{39}$$

Equation (36) implies that $(\sqrt{V_{\bar{k},s}}+\sqrt{V_{\bar{k}+1,s}})^{-2}\leq V_{\bar{k},s}^{-1}\leq e^{\kappa s}X_s$. We infer from (37) that

$$2\,\mathbb{E}[\sqrt{V_{\bar{k},s}}(\sqrt{V_{\bar{k},s}}-\sqrt{V_{\bar{k}+1,s}})]\leq e^{\kappa s}\,\mathbb{E}(X_s)\,\mathbb{E}\left[(V_{\bar{k}+1,s}-V_{\bar{k},s})^2\right] \tag{40}$$

and conclude that the lemma holds. ■

Overall,

$$\mathbb{E}\left[\left(V_{\bar{k}+1,t}-V_{\bar{k},t}\right)^2\right]\leq 2t\left[\frac{\mathbb{E}(M^2)}{b}\right]^{\bar{k}}\mathrm{Var}(M)\kappa^2\,\bar{\theta}^2+\sigma^2\,e^{-\kappa\,t}\int_0^t e^{-\kappa(t-s)}\,\mathbb{E}(X_s)\,\mathbb{E}\left[(V_{\bar{k}+1,s}-V_{\bar{k},s})^2\right]\,ds.$$

Gronwall's lemma implies that

$$\mathbb{E}\left[\left(V_{\bar{k}+1,t}-V_{\bar{k},t}\right)^2\right]\leq 2t\left[\frac{\mathbb{E}(M^2)}{b}\right]^{\bar{k}}\mathrm{Var}(M)\kappa^2\,\bar{\theta}^2\exp\left[\sigma^2\int_0^t e^{\kappa s}\mathbb{E}(X_s)ds\right].$$

We integrate this inequality over the interval $[0,T]$ and obtain:

$$\|V_{\bar{k}+1}-V_{\bar{k}}\|^2_{L^2(\Omega\times[0,T])}=\mathbb{E}\left[\int_0^T(V_{\bar{k}+1,t}-V_{\bar{k},t})^2dt\right]\leq C_T\left[\frac{\mathbb{E}(M^2)}{b}\right]^{\bar{k}}, \tag{41}$$

where $C_T=2\,\mathrm{Var}(M)\kappa^2\,\bar{\theta}^2\int_0^T t\exp\left[\sigma^2\int_0^t e^{\kappa s}\mathbb{E}(X_s)ds\right]dt$. Since $\mathbb{E}(M^2)<b$, we conclude that the sequence $V_{\bar{k},t}$ satisfies the Cauchy property and therefore has a limit in $L^2(\Omega\times[0,T])$ as $\bar{k}\to\infty$.

A.2 Convergence of the log price process

The log price process satisfies:

$$s_{\bar{k},t}=s_0+\int_0^t\left[r_s-d_s+(\alpha-1/2)V_{\bar{k},s}\right]\,ds+\int_0^t\sqrt{V_{\bar{k},s}}dW_{1,s}+J_{\bar{k},t}-\int_0^t\bar{J}_{M_{\bar{k},s}}ds. \tag{42}$$

Hence $s_{\bar{k}+1,t} - s_{\bar{k},t} = \varepsilon^{(1)}_{\bar{k}+1,t} + \varepsilon^{(2)}_{\bar{k}+1,t} + \varepsilon^{(3)}_{\bar{k}+1,t} + J_{\bar{k}+1,t} - J_{\bar{k},t}$, where

$$\begin{aligned}
\varepsilon^{(1)}_{\bar{k}+1,t} &= -\frac{1}{2}\int_0^t \left(V_{\bar{k}+1,s} - V_{\bar{k},s}\right)\, ds \\
\varepsilon^{(2)}_{\bar{k}+1,t} &= \int_0^t (\sqrt{V_{\bar{k}+1,s}} - \sqrt{V_{\bar{k},s}}) dW_{1,s} \\
\varepsilon^{(3)}_{\bar{k}+1,t} &= -\int_0^t (\bar{J}_{M_{\bar{k}+1,s}} - \bar{J}_{M_{\bar{k},s}}) ds.
\end{aligned}$$

The Cauchy-Schwarz inequality implies that

$$\mathbb{E}\left\{\left[\int_0^t (V_{\bar{k}+1,s} - V_{\bar{k},s})\, ds\right]^2\right\} \leq t\, \int_0^t \mathbb{E}\left[(V_{\bar{k}+1,s} - V_{\bar{k},s})^2\right]\, ds \leq t\, \|V_{\bar{k}+1} - V_{\bar{k}}\|^2_{L^2(\Omega\times[0,T])},$$

and therefore, by (41),

$$\left\| \varepsilon^{(1)}_{\bar{k}+1} \right\|_{L^2(\Omega\times[0,T])} \leq \sqrt{C_T}\, T \left[\frac{\mathbb{E}(M^2)}{b}\right]^{\bar{k}/2}. \tag{43}$$

Equations (39) and (40) imply that $\mathbb{E}[(\sqrt{V_{\bar{k}+1,s}} - \sqrt{V_{\bar{k},s}})^2] \leq e^{\kappa\, s}\, \mathbb{E}(X_s)\, \mathbb{E}[(V_{\bar{k}+1,s} - V_{\bar{k},s})^2]$ and therefore

$$\left\| \varepsilon^{(2)}_{\bar{k}+1} \right\|_{L^2(\Omega\times[0,T])} \leq D_T \left[\frac{\mathbb{E}(M^2)}{b}\right]^{\bar{k}/2}. \tag{44}$$

where $D_T = T\left\{2\,\mathrm{Var}(M)\kappa^2\,\bar{\theta}^2 \int_0^T e^{\kappa t}\mathbb{E}(X_t) t \exp[\sigma^2 \int_0^t e^{\kappa s}\mathbb{E}(X_s) ds] dt\right\}^{1/2}$.

The jump process is distributed as $J_{\bar{k},t} = \sum_{k=1}^{\bar{k}} j_{k,t}$, where

$$j_{k,t} = \frac{\beta(m_0 - m_1)}{\lambda_k} \sum_{n=1}^{N_{k,t}} \varepsilon_{k,t}$$

is the cumulative jump corresponding to frequency k. We note that the cumulative jump is bounded above: $|j_{k,t}| \leq \beta(m_0 - m_1)/\lambda_k$. Hence $|J_{\bar{k}+1,t} - J_{\bar{k},t}| \leq \beta(m_0 - m_1)/\lambda_{\bar{k}+1}$ and therefore

$$\|J_{\bar{k}+1} - J_{\bar{k}}\|_{L^2(\Omega\times[0,T])} \leq \frac{\sqrt{T}\beta(m_0 - m_1)}{\lambda_1\, b^{\bar{k}}}. \tag{45}$$

Let $\{m^1, \ldots, m^d\}$ denote the state space of $M_{\bar{k}+1,t}$. We note that

$$\begin{aligned}
\bar{J}_{M_{\bar{k},t}} &= \sum_{j=1}^{d} q_{i,j} \left(e^{\beta \sum_{k=1}^{\bar{k}} \lambda_k^{-1} \,\mathrm{sgn}(m^j_k - m^i_k)} - 1\right), \\
\bar{J}_{M_{\bar{k}+1,t}} &= \sum_{j=1}^{d} q_{i,j} \left(e^{\beta \sum_{k=1}^{\bar{k}+1} \lambda_k^{-1} \,\mathrm{sgn}(m^j_k - m^i_k)} - 1\right),
\end{aligned}$$

so that

$$\bar{J}_{M_{\bar{k}+1,t}} - \bar{J}_{M_{\bar{k},t}} = \sum_{j=1}^{d} q_{i,j} e^{\beta \sum_{k=1}^{\bar{k}} \lambda_k^{-1} \operatorname{sgn}(m_k^j - m_k^i)} \left(e^{\beta \lambda_{\bar{k}+1}^{-1} \operatorname{sgn}(m_k^j - m_k^i)} - 1 \right)$$

Since $e^{\beta \sum_{k=1}^{\bar{k}} \lambda_k^{-1} \operatorname{sgn}(m_k^j - m_k^i)} \leq e^{\beta \sum_{k=1}^{\infty} \lambda_k^{-1}}$ and $\left| e^{\beta \lambda_{\bar{k}+1}^{-1} \operatorname{sgn}(m_k^j - m_k^i)} - 1 \right| \leq e^{\beta \lambda_1^{-1}} \beta \lambda_{\bar{k}+1}^{-1}$, we infer that

$$|\bar{J}_{M_{\bar{k}+1,t}} - \bar{J}_{M_{\bar{k},t}}| \leq e^{2\beta \sum_{k=1}^{\infty} \lambda_k^{-1}} \beta \lambda_{\bar{k}+1}^{-1}$$

which implies that

$$\left\| \varepsilon_{\bar{k}+1}^{(3)} \right\|_{L^2(\Omega \times [0,T])} \leq \frac{e^{2\beta \lambda_1^{-1} b/(1-b)} \beta T^{3/2}}{\lambda_1 b^{\bar{k}}}. \tag{46}$$

By equations (43)–(46), the log price process $s_{\bar{k}}$ is a Cauchy sequence and therefore converges to a limiting process in $L^2(\Omega \times [0,T])$ as $\bar{k} \to \infty$.

B Risk-adjusted measure

We assign a risk premium $\Lambda(m^i) = \Lambda_i \in \mathbb{R}_{++}$ $(i = 1, ..., d)$ to each of the d Markov states and let

$$\xi_t^{(1)} = \frac{\Lambda(M_t)}{\Lambda(M_0)} \exp \left[- \int_0^t \frac{\sum_{i=1}^{d} \left(\sum_{j=1}^{d} q_{i,j} \Lambda_j \right) 1_{\{M_s = m^i\}}}{\Lambda(M_s)} ds \right]. \tag{47}$$

In order to modify the drift of the Wiener process and the jump process, we define the vector

$$\theta_t = \sqrt{V_t} \left(\alpha, - \frac{\alpha \rho + \eta_V \sigma^{-1}}{\sqrt{1 - \rho^2}} \right) + \frac{\bar{J}^*_{M_t} - \bar{J}_{M_t}}{\sqrt{V_t}} \left(1, \frac{-\rho}{\sqrt{1 - \rho^2}} \right), \tag{48}$$

where η_V is a fixed parameter and

$$\bar{J}^*_{M_t} = \sum_{i=1}^{d} 1_{\{M_t = m^i\}} \sum_{j=1}^{d} q^*_{i,j} J_{i,j} \tag{49}$$

is a modified compensator. We then assume that

$$\xi_t^{(2)} = \exp \left(- \int_0^t \theta_s dW_s - \frac{1}{2} \int_0^t \|\theta_s\|^2 ds \right). \tag{50}$$

This specification implies that the equity and volatility risk-premia are both proportional to V_t, as in Pan (2002). The market price of risk, given by $\alpha \sqrt{V_t}$, is consistent with the CAPM principle of greater risk requiring greater compensation.[14] The functional form of the volatility risk premium

[14] When the equity risk-premium is given by αdt the market price of risk is $(\alpha - r + d)/\sqrt{V_t}$ and implies that as volatility approaches its lower bound the market price of risk approaches infinity. This would potentially permit arbitrage opportunities and is the primary motivation for choosing an equity risk-premium given by $\alpha V_t dt$.

ensures that the risk neutral process remains affine.

C Proof of Proposition 2

To prove the proposition, we first need to derive the infinitesimal generator of the Skew MSM. The state of the log-price process consists of its level s_t, diffusive variance V_t, and Markov chain M_t. For notational simplicity, we denote by $X_t = (s_t, V_t)$ the vector of continuous state variables. The infinitesimal generator $\mathcal{A}^{\mathbb{Q}}$ of (X_t, M_t) is defined by

$$\mathcal{A}^{\mathbb{Q}} g\left(x, m^i\right) = \lim_{\Delta t \longrightarrow 0} \frac{\mathbb{E}^{\mathbb{Q}}\left[g\left(X_{t+\Delta t}, M_{t+\Delta t}\right) | X_t = x, M_t = m^i\right] - g\left(x, m^i\right)}{\Delta t}; \tag{51}$$

for any $(x, m^i) \in \mathbb{R} \times \mathbb{R}_+ \times \mathcal{D}$, and for any real-valued function g defined on $\mathbb{R} \times \mathbb{R}_+ \times \mathcal{D}$.

Lemma 4. *The generator $\mathcal{A}^{\mathbb{Q}}$ of (X_t, M_t) is given by*

$$\mathcal{A}^{\mathbb{Q}} g\left(x, m^i\right) = \left(q^*_{ii} + \mathcal{A}^{\mathbb{Q}}_i\right) g\left(x, m^i\right) + \sum_{j \neq i} q^*_{ij} g\left(x + J_{ij}, m^j\right), \tag{52}$$

where

$$\mathcal{A}^{\mathbb{Q}}_i g\left(x, m^i\right) = \frac{1}{2} V_t \frac{\partial^2 g}{\partial s_t^2} + \rho\sigma V_t \frac{\partial^2 g}{\partial s_t \partial V_t} + \frac{1}{2}\sigma^2 V_t \frac{\partial^2 g}{\partial V_t^2} + (r - d - \frac{V_t}{2} - \bar{J}^*_{m^i}) \frac{\partial g}{\partial s_t} + \kappa^* \left(\theta^*_{m^i} - V_t\right) \frac{\partial g}{\partial V_t}$$

is the generator of $\{X_t\}$ conditional on $M_t = m^i$ for all t.

Proof. We follow the methodology of Chourdakis (2006). We begin by considering the local variation of g under a constant regime ($M_t = m^i$ for all t). Ito's lemma implies that

$$g(X_{t+dt}, m^i) = g(X_t, m^i) + \frac{\partial g}{\partial s_t} ds_t + \frac{\partial g}{\partial V_t} dV_t + \frac{1}{2}\frac{\partial^2 g}{\partial s_t^2}(ds_t)^2 + \frac{\partial^2 g}{\partial s_t \partial V_t}(ds_t)(dV_t) + \frac{1}{2}\frac{\partial^2 g}{\partial V_t^2}(dV_t)^2.$$

Hence,

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}}\left[g(X_{t+dt}, m^i)\right] &= g(X_t, m^i) + \left(r - d - \frac{V_t}{2} - \bar{J}^*_{m^i}\right) \frac{\partial g}{\partial s_t} dt + \kappa^*(\theta^*_{m^i} - V_t) \frac{\partial g}{\partial V_t} dt \\ &\quad + \frac{1}{2} V_t \frac{\partial^2 g}{\partial s_t^2} dt + \rho\sigma V_t \frac{\partial^2 g}{\partial s_t \partial V_t} dt + \frac{1}{2}\sigma^2 V_t \frac{\partial^2 g}{\partial V_t^2} dt, \end{aligned} \tag{53}$$

and equation (52) holds.

We now consider the general case of a Markov-switching diffusion. We expand the expectation

by conditioning on the Markov state m^i:

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}}\left[g\left(X_{t+dt}, M_{t+dt}\right)|x, m^i\right] &= \sum_{j\neq i} q_{ij}^* \mathbb{E}^{\mathbb{Q}}\left[g\left(X_{t+dt}, m^j\right)|x, m^i\right] dt \\ &\quad + \left(1 + q_{ii}^* dt\right) \mathbb{E}^{\mathbb{Q}}\left[g\left(X_{t+dt}, m^i\right)|x, m^i\right] + o\left(dt\right).\end{aligned} \tag{54}$$

The first term on the right-hand side involves taking an expectation over the regime-switch and must therefore incorporate the associated jump:

$$\mathbb{E}^{\mathbb{Q}}\left[g\left(X_{t+dt}, m^j\right)|x, m^i\right] = g\left(x + J_{ij}, m^j\right) + o\left(dt\right). \tag{55}$$

The second term is given by the conditional generator

$$\mathbb{E}^{\mathbb{Q}}\left[g\left(X_{t+dt}, m^i\right)|x, m^i\right] = g\left(x, m^i\right) + \mathcal{A}_i^{\mathbb{Q}} g\left(x, m^i\right) dt + o\left(dt\right). \tag{56}$$

Substituting gives us

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}}\left[g\left(X_{t+dt}, M_{t+dt}\right)|x, m^i\right] &= \sum_{j\neq i} q_{ij}^* dt \left[g\left(x + J_{ij}, m^j\right) + \mathcal{A}_j^{\mathbb{Q}} g\left(x, m^j\right) dt\right] \\ &\quad + \left(1 + q_{ii}^* dt\right) \left[g\left(x, m^i\right) + \mathcal{A}_i^{\mathbb{Q}} g\left(x, m^i\right) dt\right] + o\left(dt\right).\end{aligned} \tag{57}$$

We conclude that the lemma holds. ■

We observe that $\mathcal{A}_i^{\mathbb{Q}} g\left(x, m^i\right)$ is the generator of a standard stochastic volatility model.

Equipped with this result, we can now proceed to deriving the characteristic function. Before doing so, we restate a well-known result for the constant regime case below.

Lemma 5 (**Characteristic function under a constant regime**). *The characteristic function of s_T under a constant regime m^i,*

$$\varphi_i(u, \tau; s, V) = \mathbb{E}^{\mathbb{Q}}\Big(e^{iu s_T}\Big| \, s_t = s, V_t = V, \, M_{t'} = m^i \text{ for all } t' \in [t, T]\Big) \tag{58}$$

satisfies

$$\varphi_i\left(u, \tau\right) = e^{\mathrm{i}us} e^{C_i(u,\tau) + D(u,\tau)\, V}, \tag{59}$$

where the complex coefficients $C_i(u, \tau)$ and $D(u, \tau)$ are given by (18)–(21).

The functions $D(u, \tau)$ and $C_i(u, \tau)$ satisfy a system of ordinary differential equations, which we

obtain by applying the conditional generator $\mathcal{A}_i^{\mathbb{Q}}$ to the function $\varphi_i(u,\tau)$:

$$\begin{aligned}\mathcal{A}_i^{\mathbb{Q}}\varphi_i\left(u,\tau;s,V\right) &= \left[-\frac{u^2V}{2}+\mathrm{i}u\rho V\sigma D(u,\tau)+\frac{\sigma^2V}{2}D(u,\tau)^2\right]\varphi_i \\ &\quad+\left[\mathrm{i}u(r-d-\frac{V}{2}-\bar{J}^*_{m^i})+\kappa^*(\theta^*_{m^i}-V)D(u,\tau)\right]\varphi_i.\end{aligned} \tag{60}$$

The time derivative of $\varphi_i(u,\tau)$ is

$$\frac{\partial\varphi_i}{\partial\tau}\left(u,\tau\ s,V\right)=\left[\dot{C}_i(u,\tau)+\dot{D}(u,\tau)\,V\right]\varphi_i\left(u,\tau\ s,V\right),$$

where $\dot{C}_i(u,\tau)\equiv\partial C_i/\partial\tau$ and $\dot{D}(\tau)\equiv\partial D(u,\tau)/\partial\tau$. Since the characteristic function is a martingale, we know that

$$\mathcal{A}_i^{\mathbb{Q}}\varphi_i\left(u,\tau\ s,V\right)=\frac{\partial\varphi_i}{\partial\tau}\left(u,\tau;s,V\right).$$

Hence,

$$\dot{D}(u,\tau)=-\frac{u^2}{2}+\mathrm{i}u\rho\sigma D(u,\tau)+\frac{\sigma^2}{2}D(u,\tau)^2-\frac{\mathrm{i}u}{2}-\kappa^*D(u,\tau), \tag{61}$$

$$\dot{C}_i(u,\tau)=\mathrm{i}u\left(r-d-\bar{J}^*_{m^i}\right)+\kappa^*\theta^*_{m^i}D(u,\tau), \tag{62}$$

which can also be checked directly from the function forms of D and C_i.

We next consider Skew MSM with regime-switches, and consider the conditional characteristic function:

$$\psi_i(u,\tau;s,V)=\mathbb{E}^{\mathbb{Q}}\Big(e^{\mathrm{i}us_T}\Big|\,s_t=s,V_t=V,\,M_t=m^i\Big). \tag{63}$$

We observe that $\psi_i(u,\tau;s,V)=e^{\mathrm{i}us}\mathbb{E}^{\mathbb{Q}}\Big(e^{\mathrm{i}u(s_T-s)}\Big|\,s_t=s,V_t=V,\,M_t=m^i\Big)$. An educated guess is that

$$\psi_i(u,\tau;s,V)=e^{\mathrm{i}us+D(u,\tau)V}\bar{\psi}_i(u,\tau) \tag{64}$$

for all $i\in\{1,\ldots,d\}$. Applying the generator in equation (52), and exploiting the exponential affine structure as well as the martingale property of the characteristic function, we get

$$\frac{\partial\psi_i}{\partial\tau}(u,\tau;s,V)=\mathcal{A}^{\mathbb{Q}}\psi_i\left(u,\tau;;s,V\right)=\left(q^*_{ii}+\mathcal{A}_i^{\mathbb{Q}}\right)\psi_i\left(u,\tau;s,V\right)+\sum_{j\neq i}q^*_{ij}\psi_j(u,\tau;s,V)e^{\mathrm{i}uJ_{ij}}. \tag{65}$$

We observe that

$$\begin{aligned}\mathcal{A}_i^{\mathbb{Q}}\psi_i &= \left[-\frac{u^2V}{2}+\mathrm{i}u\rho V\sigma D(u,\tau)+\frac{\sigma^2V}{2}D(u,\tau)^2+\mathrm{i}u(r-d-\frac{V}{2}-\bar{J}^*_{m^i})+\kappa^*(\theta^*_{m^i}-V)D(u,\tau)\right]\psi_i \\ &= \left[\dot{C}_i(u,\tau)+\dot{D}(u,\tau)\,V\right]\psi_i\end{aligned}$$

Hence,

$$\frac{\partial \psi_i}{\partial \tau}(u,\tau;s,V) = \left[q^*_{ii} + \dot{C}_i(u,\tau) + \dot{D}(u,\tau)\,V\right]\psi_i\,(u,\tau;s,V) + \sum_{j\neq i} q^*_{ij}\psi_j(u,\tau;s,V)e^{\mathrm{i}uJ_{ij}}.$$

Since $\psi_i(u,\tau;\,s,V) = e^{\mathrm{i}us+D(u,\tau)\,V}\bar{\psi}_i(u,\tau)$, we infer that

$$\frac{\partial \bar{\psi}_i}{\partial \tau}(u,\tau) + \dot{D}(u,\tau)\,V\,\bar{\psi}_i(u,\tau) = \left[q^*_{ii} + \dot{C}_i(u,\tau) + \dot{D}(u,\tau)\,V\right]\bar{\psi}_i\,(u,\tau) + \sum_{j\neq i} q^*_{ij}\bar{\psi}_j(u,\tau)e^{\mathrm{i}uJ_{ij}}$$

and conclude that the proposition holds.

D Filtering

D.1 Particle filter for stock returns

By rearranging equation (26) and substituting into equation (27), we can remove $\varepsilon_{1,t}$ from the volatility transition equation:

$$v_{t+1} = v_t + \kappa\left(\theta_{M_{t+1}} - v_t\right)\Delta t + \rho\sigma\left[s_t - s_{t-1} - \mu_t\Delta t - \hat{J}_t\left(M_t, M_{t-1}\right)\right] + \sqrt{1-\rho^2}\sigma\sqrt{v_t\Delta t}\varepsilon_{2,t}.$$

The particle filter is then constructed as follows.

Step 1 (Propagation): We propagate the particles forward:

$$\begin{aligned}
\hat{M}^{(k)}_{t+1} &\sim \mathbb{P}(M_{t+1}\Big|M^{(k)}_t), \\
\hat{v}^{(k)}_{t+1} &= v^{(k)}_t + \kappa(\theta_{\hat{M}^{(k)}_{t+1}} - v_t)\Delta t + \rho\sigma\left[s_t - s_{t-1} - \mu_t\Delta t - \hat{J}_t\left(M^{(k)}_t, M^{(k)}_{t-1}\right)\right] \\
&\quad + \sqrt{1-\rho^2}\sigma\sqrt{v_t\Delta t}\varepsilon^{(k)}_{2,t}.
\end{aligned}$$

The set of particles $\left(\hat{v}^{(k)}_{t+1}, \hat{M}^{(k)}_{t+1}, M^{(k)}_t\right)_{k=1,\ldots,K}$ targets the distribution of (v_{t+1}, M_{t+1}, M_t) conditional on $y_{1:t}$.

Step 2 (Importance weights): We assign to each particle a weight equal to its observation density: $w^{(k)}_{t+1} = f(y_{t+1}|\hat{v}^{(k)}_{t+1}, \hat{M}^{(k)}_{t+1}, M^{(k)}_t)$. The weights are then normalized

$$\pi^{(k)}_{t+1} = \frac{w^{(k)}_{t+1}}{\sum_{k=1}^K w^{(k)}_{t+1}}. \tag{66}$$

The particles and associated normalized weights $\left[\left(\hat{v}_{t+1}^{(k)}, \hat{M}_{t+1}^{(k)}, M_t^{(k)}\right); \pi_{t+1}^{(k)}\right]_{k=1,\ldots,K}$ target the posterior density of (v_{t+1}, M_{t+1}, M_t) given $y_{1:t}$.

Step 3 (Resampling): In the final step, we uniformly resample from the approximate posterior distribution. We draw $(v_{t+1}^{(1)}, M_{t+1}^{(1)}, M_t^{(1)})$ from $\left[(\hat{v}_{t+1}^{(k)}, \hat{M}_{t+1}^{(k)}, M_t^{(k)}); \pi_{t+1}^{(k)}\right]_{k=1,\ldots,K}$. We repeat this operation for $k' = 1, \ldots, k$ and obtain the date$-t$ filter. [15]

D.2 Particle filter for stock and option data

Given the particle filter at date t, the construction of the period $t+1-$filter proceeds in four steps.

Step 1: For each k, we simulate d particles forward as follows. We let $\hat{M}_{t+1}^{(k,j)} = m^j$, simulate $\hat{v}_{t+1}^{(k,j)}$ from the distribution of v_{t+1} conditional on $(M_{t-1}^{(k)}, M_t^{(k)}, \hat{M}_{t+1}^{(k,j)} = m^j, v_t^{(k)})$,[16] and assign to the particle a weight $w_{t+1}^{(k,j)} = \mathbb{P}[M_{t+1} = m^j | M_t^{(k)} = m^i]$. This produces a set of $d \times K$ weighted particles.

Step 2: For each $j \in \{1, \ldots, d\}$, we resample R particles from

$$(\hat{v}_{t+1}^{(k,j)}, m^j, M_t^{(k)}), \qquad k = 1, \ldots, K,$$

with probabilities $w_{t+1}^{(k,j)} / \sum_{k'=1}^{K} w_{t+1}^{(k',j)}$. The resulting set consists of

$$(\hat{\hat{v}}_{t+1}^{(r,j)}, m^j, \hat{\hat{M}}_t^{(r,j)}), \tag{67}$$

where $r \in \{1, \ldots, R\}$ and $j \in \{1, \ldots, d\}$.

Step 3: We use the observation equation to assign a new weight to each of the $R \times d$ particles: $\hat{\hat{w}}_{t+1}^{(r,j)} = \mathbb{P}(M_{t+1} = m^j | \hat{\hat{M}}_t^{(r,j)}) f\left(y_{t+1} \middle| \hat{\hat{v}}_{t+1}^{(r,j)}, M_{t+1} = m^j, \hat{\hat{M}}_t^{(r,j)}\right)$, and define the normalized weights

$$\pi_{t+1}^{(r,j)} = \frac{\hat{\hat{w}}_{t+1}^{(r,j)}}{\sum_{r'=1}^{R} \sum_{j'=1}^{d} \hat{\hat{w}}_{t+1}^{(r',j')}}. \tag{68}$$

Step 4: We resample K particles from $(\hat{\hat{v}}_{t+1}^{(r,j)}, m^j, \hat{\hat{M}}_t^{(r,j)})$ with probabilities $\pi_{t+1}^{(r,j)}$. The final set of resampled particles $\left(v_{t+1}^{(k)}, M_{t+1}^{(k)}, M_t^{(k)}\right)_{k=1,\ldots,K}$ is now equally weighted.

[15] This step is necessary to ensure that the sample does not disperse over time and, in addition, that the particles characterize the posterior distribution efficiently.

[16] Recall that $\hat{v}_{t+1}^{(k,j)} \sim \mathcal{N}\left[v_t^{(k)} + \kappa(\theta_{\hat{M}_{t+1}^{(k,j)}} - v_t)\Delta t + \rho\sigma[s_t - s_{t-1} - \mu_t \Delta t - \hat{J}_t(M_t^{(k)}, M_{t-1}^{(k)})]; \left(1-\rho^2\right)\sigma^2 v_t \Delta t\right]$.

The first two steps achieve a stratified propagation of the filter based only the information $y_{1:t}$ available up to date t. In Step 1, we propagate for every $j \in \{1, .., d\}$ the v_t particles K times forward conditional on $M_{t+1} = m^j$. This procedure generates a set of $d \times K$ particles. In Step 2, for every $j = 1, .., d$, we resample R particles conditional on $M_{t+1} = m^j$, thus producing a set of $d \times R$ particles. Each M_{t+1} state is now represented by R particles. By choosing $R < K$ we can reduce the number of evaluations of the observation density in subsequent stages.

The third and fourth steps update the filter by taking into account the new measurement y_{t+1}. They proceed in the same way as the previous particle filter method. In the Step 3, a new weight is assigned to each particle by evaluating the observation density, which is now a function of both returns and option prices. Note that this step requires only $d \times R$ evaluations of the observation density (instead of K evaluations under the standard filter). In Step 4, K particles are resampled from the set of $d \times R$ particles.

The particle filter can be computed recursively at dates $t = 1, \ldots, T$. We then estimate the likelihood of $y_{1:T}$ by

$$\hat{\mathcal{L}} = \prod_{t=1}^{T} \left[\frac{1}{Rd} \sum_{r=1}^{R} \sum_{j=1}^{d} \hat{\hat{w}}_t^{(r,j)} \right]. \tag{69}$$

E Maximum likelihood estimation of alternative models

MSM, MSM-J, and GARCH-type models all have closed form likelihood functions and are simple to estimate. In contrast, the SV, NA-SV, SVJ0, and SVJ1 models have no closed-form likelihood. For their estimation, we use the particle filter as detailed in Christoffersen, Jacobs and Mimouni (2010).

Table E.1: In-sample maximum likelihood estimation of alternative models

A. Jump diffusion specifications

	α	κ	θ	σ	ρ	λ	μ_J	σ_J
SV	4.34	5.72	0.018	0.34	-0.57	-	-	-
NA-SV	3.88	2.28	0.020	2.38	-0.58	-	-	-
SVJ0	4.37	3.43	0.012	0.22	-0.56	0.46	-0.044	0.058
SVJ1	4.71	4.02	0.016	0.25	-0.61	55.8	-0.016	0.045

B. GARCH specifications

	μ	ω	α	γ	β
GARCH	0.04	6.29E-7	0.079	0.092	-
GJR-GARCH	0.04	7.93E-7	0.029	0.086	0.926
ZARCH	0.04	1.44E-4	0.032	0.092	0.919

The table reports the ML estimation results for the parameters of the alternative models using daily S&P 500 log excess returns from March 6, 1957 to September 28, 2007. In Panel A, we report the parameter estimates for the jump diffusion specifications. In Panel B, we report the estimates for our GARCH specifications.

References

Adrian, T. and Rosenberg, J.: 2008, Stock returns and volatility: Pricing the short-run and long-run components of market risk, *Journal of Finance* **63**, 2997–3030.

Andersen, T., Fusari, N. and Todorov, V.: 2012, Parametric inference and dynamic state recovery from option panels, *Technical report*, Northwestern University.

Andersen, T. and Sorensen, B. E.: 1996, GMM estimation of a stochastic volatility model: A monte carlo study, *Business and Economic Statistics* **14**, 328–352.

Bakshi, G., Cao, C. and Chen, Z.: 1997, Empirical performance of alternative option pricing models, *Journal of Finance* **52**, 2003–2049.

Bates, D.: 1996, Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options, *Review of Financial Studies* **9**, 69–107.

Bates, D.: 2000, Post - '87 crash fears in the S&P 500 futures option market, *Journal of Finance* **94**, 181–238.

Bates, D.: 2006, Maximum likelihood estimation of latent affine processes, *Review of Financial Studies* **19**, 909–965.

Bates, D.: 2012, U.S. stock market crash risk, 1926-2010, *Journal of Financial Economics* **105**, 229–259.

Broadie, M., Chernov, M. and Johannes, M.: 2007, Model specification and risk premia: Evidence from futures options, *Journal of Finance* **62**, 1453–1490.

Calvet, L. E. and Fisher, A. J.: 2001, Forecasting multifractal volatility, *Journal of Econometrics* **105**, 27–58.

Calvet, L. E. and Fisher, A. J.: 2004, How to forecast long-run volatility: Regime switching and the estimation of multifractal processes, *Journal of Financial Econometrics* **2**, 49–83.

Calvet, L. E., Fisher, A. J. and Mandelbrot, B. B.: 1997, Cowles foundation working papers 1164-1166, *Technical report*, Yale University.

Calvet, L. E., Fisher, A. J. and Thompson, S.: 2006, Volatility comovement: A multifrequency approach, *Journal of Econometrics* **131**, 179–215.

Calvet, L. E., Fisher, A. J. and Wu, L.: 2011, Dimension-invariant dynamic term structures, *Technical report*, HEC Paris.

Calvet, L. and Fisher, A. J.: 2008a, *Multifractal Volatility: Theory, Forecasting and Pricing*, Elsevier - Academic Press.

Calvet, L. and Fisher, A. J.: 2008b, Multifrequency jump-diffusions: An equilibrium approach, *Journal of Mathematical Economics* **44**(2), 207 – 226.

Campbell, J. Y., Giglio, S., Polk, C. and Turley, R.: 2012, An intertemporal CAPM with stochastic volatility, *Technical report*, Harvard University.

Carr, P. and Madan, D.: 1999, Option valuation using the fast Fourier transform, *J. Comp. Finance* **2**, 61–73.

Carr, P. P., Geman, H., Madan, D. B. and Yor, M.: 2002, The fine structure of asset returns: An empirical investigation, *Journal of Business* **75**, 305–332.

Carr, P. P. and Wu, L.: 2004, Time changed levy processes and option pricing, *Journal of Financial Economics* **71**, 113–141.

Chernov, M., Gallant, R. A., Ghysels, E. and Tauchen, G.: 2003, Alternative models for stock price dynamics, *Journal of Econometrics* **116**(1-2), 225–257.

Chernov, M. and Ghysels, E.: 2000, A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of option valuation, *Journal of Financial Economics* **56**, 407–458.

Chourdakis, K.: 2006, Switching Lévy models in continuous time: Finite distributions and option pricing, *Technical report*, University of Essex, Centre for Computational Finance and Economic Agents (CCFEA) Working Paper.

Christoffersen, P., Dorion, C., Jacobs, K. and Wang, Y.: 2010, Volatility components: Affine restrictions and non-normal innovations, *Journal of Business and Economic Statistics* **28**, 483–502.

Christoffersen, P., Heston, S. and Jacobs, K.: 2009, The shape and term-structure of the index option smirk: Why multifactor stochastic volatility models work so well, *Management Science* **55**, 1914–1932.

Christoffersen, P., Jacobs, K. and Mimouni, K.: 2010, Volatility dynamics for the S&P 500: Evidence from realized volatility, daily returns, and option prices, *Review of Financial Studies* **23**, 3141–3189.

Christoffersen, P., Jacobs, K., Ornthanalai, C. and Wang, Y.: 2008, Option valuation with long-run and short-run volatility components, *Journal of Financial Economics* **90**, 272–297.

Corsi, F.: 2009, A simple approximate long-memory model of realized volatility, *Journal of Financial Econometrics* **7**, 174–196.

Corsi, F., Fusari, N. and Vecchia, D. L.: 2013, Realizing smiles: Option pricing with realized volatility, forthcoming, *Journal of Financial Economics*.

Doucet, A. and Johansen, A. M.: 2011, A tutorial on particle filtering and smoothing: fifteen years later, *Technical report*, Oxford Handbook of Nonlinear Filtering.

Duffie, D., Pan, J. and Singleton, K.: 2000, Transform analysis and asset pricing for affine jump-diffusions, *Econometrica* **68**, 1343–1376.

Durham, G. and Park, Y.: 2012, Beyond stochastic volatility and jumps in returns in volatility, forthcoming, *Journal of Business and Economic Statistics*.

Egloff, D., Leippold, M. and Wu, L.: 2010, The term structure of variance swap rates and optimal variance swap investments, *Journal of Financial and Quantitative Analysis* **45**, 1279–1310.

Elliott, R., Siu, T. and Badescu, A.: 2011, On pricing and hedging options in regime-switching models with feedback effect, *Journal of Economic Dynamics and Control* **35**, 694–713.

Elliott, R., Siu, T. and Chan, L.: 2007, Pricing volatility swaps under Heston's stochastic volatility model with regime switching, *Applied Mathematical Finance* **14**, 41–62.

Elliott, R., Siu, T., Chan, L. and Lau, J.: 2007, Pricing options under a generalized Markov-modulated jump-diffusion model, *Stochastic Analysis and Applications* **25**, 821–843.

Engle, R. F. and Lee, G.: 1999, A permanent and transitory component model of stock return volatility, *Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W.J. Granger*, pp. 475–497.

Fang, F. and Oosterlee, C.: 2008, A novel pricing method for european options based on fourier-cosine series expansions, *SIAM Journal on Scientific Computing* **31**, 826–848.

Garcia, R., Ghysels, E. and Renault, E.: 2009, The econometrics of option pricing, *in* Y. Ait-Sahalia and L. P. Hansen (eds), *Handbook of Financial Econometrics*, Elsevier, pp. 479–452.

Glosten, L. R., Jagannathan, R. and Runkle, D. E.: 1993, On the relation between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance* **48**(5), 1779–1801.

Gordon, N. J., Salmond, D. J. and Smith, A. F. M.: 1993, A novel approach to nonlinear/non-gaussian bayesian state estimation, *IEE-Proceedings* **140**, 107–133.

Heston, S.: 1993, A closed-form solution for options with stochastic volatility with applications to bond and currency options, *Review of Financial Studies* **6**, 327–343.

Hull, J. and White, A.: 1987, The pricing of options on assets with stochastic volatility, *Journal of Finance* **42**(2), 281–300.

Ikeda, N. and Watanabe, S.: 1977, A comparison theorem for solutions of stochastic differential equations and its applications, *Osaka Journal of Mathematics* **14**, 619–633.

Johannes, M., Polson, N. and Stroud, J.: 2009, Optimal filtering of jump-diffusions: Extracting latent states from asset prices, *Review of Financial Studies* **22**, 2759–2799.

Jones, C.: 2003, The dynamics of stochastic volatility: Evidence from underlying and options markets, *Journal of Econometrics* **116**, 181–224.

Kaeck, A. and Alexander, C.: 2012, Volatility dynamics for the S&P 500: Further evidence from non-affine, multi-factor jump diffusions, *Journal of Banking and Finance* **36**(11), 3110 – 3121.

Kitagawa, G.: 1996, Monte carlo filter and smoother for non-gaussian non-linear state space models, *Computational and Graphics Statistics* **5**, 1–25.

Li, G. and Zhang, C.: 2013, Diagnosing affine models of option pricing: Evidence from vix, *Journal of Financial Economics* **107**, 199–219.

Liu, J., Pan, J. and Wang, T.: 2005, An equilibrium model of rare-event premia and its implication for option smirks, *Review of Financial Studies* **18**, 131–164.

Lord, R., Koekkoek, R. and van Dijk, D.: 2010, A comparison of biased simulation schemes for stochastic volatility models, *Quantitative Finance* **10**(2), 177–194.

Lux, T.: 2008, The markov-switching multifractal model of asset returns: GMM estimation and linear forecasting of volatility, *Journal of Business and Economic Statistics* **26**, 194–210.

Lux, T. and Kaizoji, T.: 2007, Forecasting volatility and volume in the tokyo stock market: Long memory, fractality, and regime switching, *Journal of Economic Dynamics and Control* **31**, 1808–1843.

Malik, S. and Pitt, M. K.: 2011, Modeling stochastic volatility with leverage and jumps: A simulated maximum likelihood approach via particle filtering, *Technical report*, Banque de France.

Merton, R. C.: 1976, Option pricing when the underlying stock returns are discontinuous, *Journal of Financial Economics* **3**, 125–144.

Merton, R. C.: 1980, On estimating the expected return on the market: An exploratory analysis, *Journal of Financial Economics* **8**, 323–361.

Murphy, K. M. and Topel, R. H.: 1985, Estimation and inference in two-step econometric models, *Journal of Business and Economic Statistics* **3**, 370–379.

Naik, V. and Lee, M.: 1990, General equilibrium pricing of options on the market portfolio with discontinuous returns, *Review of Financial Studies* **3**, 493–521.

Øksendal, B.: 2010, *Stochastic Differential Equations: An Introduction with Applications*, Springer.

Pan, J.: 2002, The jump-risk premia implicit in options: Evidence from an integrated time-series study, *Journal of Financial Economics* **63**, 3–50.

Pitt, M. and Shephard, N.: 1999, Filtering via simulation: Auxiliary particle filters, *American Statistical Association* **94**, 590–599.

Vuong, Q.: 1989, Likelihood ratio tests for model selection and non-nested hypotheses, *Econometrica* **57**, 307–333.

Zakoian, J.-M.: 1994, Threshold heteroskedastic models, *Journal of Economic Dynamics and Control* **18**(5), 931–955.